The Performance-Learning Gap in the Use of Generative AI for English as a Foreign Language Learning

Ryoma Matsumoto[1*], Nabil Hasan Saleh Al-Kumaim [2]

[1,2] Universiti Teknologi Malaysia, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101100356

Received: 31 October 2026; Accepted: 06 November 2026; Published: 26 November 2026

ABSTRACT

Generative AI can enhance writing and related text-based tasks in English as a Foreign Language (EFL) learning. However, using AI does not necessarily build lasting language skills. This conceptual paper discusses the performance-learning gap between AI-assisted output and learners' independent performance. It refers to what students can produce with AI support and what they can later perform independently. From an Information Systems view, this gap means visible output from human-AI teamwork may not accurately reflect learners' independent ability. Drawing on the Theory of Effective Use, Task-Technology Fit, socio-technical configuration, and knowledge-management concepts such as the SECI model, the paper argues that AI may support immediate performance goals without necessarily supporting durable learning. Methodologically, this paper adopts a conceptual literature review and theory-synthesis approach, integrating insights from Information Systems, knowledge management, cognitive psychology, and second language acquisition. The resulting input-process-output-feedback framework explains how socio-technical conditions and AI use patterns may widen or narrow the performance-learning gap. The gap widens when AI eases task completion but hinders students from noticing issues, explaining processes, evaluating work, and reconstructing knowledge. Conversely, if teachers design instruction carefully, govern AI use, enforce process visibility, and test without AI, the gap might narrow.

Keywords: Generative AI; English as a Foreign Language (EFL) learning; performance-learning gap; effective use; Task-Technology Fit; socio-technical configuration; outcome ambiguity; knowledge retention

INTRODUCTION

Generative AI is rapidly emerging as a critical support tool in EFL learning. Students use it for brainstorming, correcting grammatical errors, translating, polishing their writing, getting feedback, and even for example sentences. This technology lowers the barriers of language creation and supports students in ways typical school settings might struggle to offer constantly. Current academic discourse highlights AI's possible help in education as well as its usefulness in crafting content, giving feedback, and tailoring an experience for individuals. Still, there is an emphasis on the necessity of rules and guidelines aimed at humans (Kasneci et al., 2023; Miao et al., 2023). Furthermore, existing literature suggests that AI-enabled writing tools can support language learning when integrated appropriately into writing instruction and mediated by teachers (Godwin-Jones, 2022). Recent work has also identified both the affordances of ChatGPT for language learning and the digital competencies required for its ethical and effective use (Kohnke et al., 2023).

The central issue here, however, extends beyond determining whether AI is inherently beneficial or detrimental. Instead, a more pressing dilemma involves how to interpret student work when they have had major AI assistance. Traditionally, a fluent paragraph or correct translation signalled successful language acquisition. Yet with AI, such achievement could mean the tool contributed the majority of the output. This creates severe outcome ambiguity; the output appears polished but questions remain about the true scope of what the student actually learned.

This paper conceptualises this as "the performance-learning gap": the divergence between what a learner achieves with AI and what they truly understand on their own. For instance, learners might produce highly sophisticated writing with AI assistance, only to struggle significantly without it. This does not mean that AI use prevents learning; rather, it highlights that improved output should not be equated with learning because better output does not necessarily mean that the learner has fully grasped the material.

Complicating this dynamic further, recent evidence indicates that individuals can receive immediate writing support from AI and improve task performance without corresponding significant gains in knowledge or transfer (Fan et al., 2025).

This situation is especially important in information systems research, since AI redefines the view of the connection between technology use and task completion, and personal skill growth. The Theory of Effective Use emphasises moving beyond simply adopting a tool to actually achieving worthwhile ends (Burton-Jones & Grange, 2013). Similarly, in EFL learning, the aim extends beyond task completion; the goal is to develop abilities that students retain long after the AI aid stops.

As Goodhue and Thompson (1995) originally established, information technology adds value when it is utilised properly and inherently fits the parameters of the supported task. Extending this logic to generative AI, a tool perfectly calibrated to execute an immediate task does not automatically guarantee the formation of stable, durable capabilities once AI support is removed.

Applying a socio-technical perspective to EFL learning, the ways that AI influences learning may depend on variables such as the structure of assignments, assessment strategies, teacher guidance, AI governance, and institutional norms (Baxter & Sommerville, 2011; Orlikowski, 2000).

Ultimately, this conceptual paper aims to elucidate the performance-learning gap brought on by AI in EFL contexts. Further, it aims to develop a framework showing scenarios where the gap might widen or narrow. The present analysis focuses primarily on writing and related text-based EFL tasks, where AI-assisted output can be compared with learners' independent performance. Its application to speaking and listening remains a subject for future research.

This conceptual paper makes three main contributions. First, it clarifies the performance-learning gap as a conceptual problem in generative AI-assisted EFL learning, especially where visible task performance can be mistaken for durable learner ability. Second, it brings this educational issue into Information Systems by explaining the gap through effective use, Task-Technology Fit, and socio-technical configuration. Third, it proposes an initial framework for configuring AI-assisted EFL learning so that students do not merely complete tasks with AI, but are also required to notice, explain, evaluate, retrieve, and reconstruct language knowledge. In this sense, the paper contributes not by presenting empirical findings, but by offering a conceptual basis for later empirical research and practical teacher guidance.

LITERATURE REVIEW ON GENERATIVE AI-ASSISTED EFL LEARNING

This paper employs a purposive conceptual literature review rather than a systematic review. The literature was drawn primarily from a pre-existing collection of academic sources compiled for the first author's doctoral research on generative AI-assisted EFL learning. Sources were selected for their relevance to technology use, task fit, socio-technical conditions, knowledge internalisation, generative AI-assisted EFL learning, and learning mechanisms. Foundational theoretical works were synthesised alongside recent empirical studies by mapping their contributions to four framework components: socio-technical conditions, AI use patterns, learning mechanisms, and learning outcomes. This synthesis provides the basis for the conceptual framework and propositions without claiming exhaustive coverage of the literature.

Generative AI has become an influential innovation in modern educational practice. Its significance in English as a Foreign Language (EFL) learning stems from the text-based nature of language tasks and the difficulty some learners have completing them independently. Generative AI can be used to write sentences, correct grammar, translate thoughts, rephrase, give examples, and provide explanations. This tool may be especially useful in EFL

learning where students might lack confidence, vocabulary, grammatical awareness, or individual teacher feedback. Existing literature discusses the potential of AI to contribute to education through interaction, feedback, and personalised learning, as well as the need for human-centred guidance and responsible usage (Godwin-Jones, 2022; Kasneci et al., 2023; Miao et al., 2023).

Generative AI offers immediate feedback, flexible practice, and repeated explanations that may be valuable in EFL learning (Godwin-Jones, 2022; Kasneci et al., 2023). Recent studies report improvements in EFL writing following ChatGPT feedback, while learners also describe concerns about its use (Polakova & Ivenz, 2024; Teng, 2024).

At the same time, the usefulness of generative AI in education cannot be evaluated based only on convenience. A tool that helps students complete a task efficiently does not necessarily ensure durable knowledge construction. This aspect is especially relevant in EFL learning where language skills include not only the ability to write a seemingly correct text but also the abilities to remember vocabulary, understand grammar, choose appropriate expressions, organise meaning, and apply language knowledge in new situations. According to recent research, generative AI could enhance task performance while also altering students' self-regulated learning processes and potentially encouraging technological dependence or metacognitive laziness (Fan et al., 2025). Structured LLM-supported strategy instruction has likewise been associated with improvements in EFL writing performance, self-regulated learning strategies, and motivation, indicating that learning outcomes depend partly on how AI is embedded in instruction (Liu et al., 2026). Thus, success with the aid of AI should be evaluated carefully, especially considering that the final output could be used as evidence of students' competence.

AI-supported writing is a good example of this problem. Writing assignments involve planning meaning, selecting vocabulary, writing sentences, detecting mistakes, revising drafts, and evaluating whether the final output communicates what was intended. All of these steps not only comprise the process of completing the task but are also elements of the learning process itself. If AI works as a partner in this process, it could support these learning processes. For example, the student might compare his or her original sentence with the AI-suggested version, ask for an explanation of the correction, and then rewrite the sentence independently. In such cases, AI can assist noticing, explaining, evaluating, and reconstructing. Recent research comparing teacher- and ChatGPT-generated feedback further suggests that the educational value of AI feedback is shaped by how students engage with and incorporate it into revision (Zou et al., 2026). Such activities relate to second language acquisition theories, which assume conscious attention to linguistic structures and the importance of output for generating learning opportunities (Schmidt, 1990; Swain, 1995; Swain & Lapkin, 1995).

However, the same tool could weaken learning if it becomes a substitute for production. If the student asks AI to write a paragraph, correct all mistakes, and polish the output, he or she might complete the assignment but do it without sufficient internal processing. The student might submit a fluent text but not be able to explain why certain grammatical choices were made, repeat similar expressions later, or apply knowledge to a new topic. Such a situation relates to the notion of cognitive offloading, which means the outsourcing of mental effort to an external tool in order to minimise internal cognitive work (Risko & Gilbert, 2016). Cognitive offloading is not necessarily negative but, in the case of EFL learning, it may become problematic when learners offload the very processes through which language ability is supposed to develop.

Such a problem goes far beyond cheating or students' laziness because it is also related to how assignments and assessments are designed. When the assignment is focused on the quality of output, students may have an incentive to use AI in such a way that they can maximise visible performance. A polished paragraph might be the result of the student's efforts, the efforts of AI, or both. Therefore, the final output becomes an ambiguous measure of the student's independent ability.

Such a risk becomes even more obvious when learning is evaluated after the completion of the task. Durable learning involves memory retention, reuse, and transfer. Retrieval-practice research shows that memory retention is enhanced by requiring learners to retrieve information rather than receive or recognise it (Roediger & Karpicke, 2006). At the same time, transfer research suggests that learning should be examined by considering whether knowledge can be applied beyond the original learning situation and under what conditions such transfer occurs (Barnett & Ceci, 2002).

However, there is still a need to make an explicit connection between these aspects and the difference between immediate AI-assisted performance and durable learner capability. This issue is especially important in EFL learning because the ultimate goal is not merely to create English text but to develop the learner's ability to use English independently.

Taken together, these studies indicate the need to distinguish immediate AI-assisted performance from durable learner capability. The following section develops this distinction more explicitly as an Information Systems problem.

From AI-Assisted Performance to Durable Learning

The performance-learning gap arises when a distinction is made between AI-assisted task execution and the actual learning that remains durable within a student. When an algorithmic system autonomously generates a flawless paragraph or perfect sentence, this can be understood as AI-assisted task performance. On the other hand, durable learning is about a student being able to retain and reuse the relevant linguistic structures independently later on, particularly when AI does not do it for them. While these two might occasionally overlap, treating them as one and the same would be wrong.

The performance-learning gap is distinct from cognitive offloading, the illusion of competence, and learning transfer. It refers to the discrepancy between AI-assisted task performance and the learner's independently retained capability. Cognitive offloading is a process of delegating mental work to an external tool and may contribute to widening the gap (Risko & Gilbert, 2016). The illusion of competence concerns an inaccurate judgement of one's own knowledge or ability (Koriat & Bjork, 2005). In contrast, the performance-learning gap can exist even when learners accurately recognise their limitations. Learning transfer is not the gap itself but an indicator of whether knowledge can be applied beyond the original task (Barnett & Ceci, 2002). Thus, the gap is the central phenomenon, offloading is a possible contributing process, perceived competence concerns self-assessment, and transfer provides evidence of durable learning.

This becomes particularly important in EFL settings, where tasks often serve two goals. First, there is the obvious need to produce something - such as writing an essay or doing a translation. Second, completing these tasks should help students build language skills by experimenting with new vocabulary or grammar, logically organising meaning, or noticing linguistic deficits in their knowledge. For instance, writing a story is not merely about completing it, but retrieving vocabulary and testing syntactic forms along the way.

Generative AI possesses the capacity to facilitate this dual mandate – it acts as a genuine catalyst when students use it to compare different ways of saying something, interrogate the rationale behind a correction, evaluate which suggestions work best, or reconstruct responses on their own terms afterwards. However, problems emerge when students rely on AI to generate, repair, or superficially polish output without cognitively processing the language themselves. Consequently, even though the assignment is completed, the underlying educational function is severely weakened.

The underlying danger is clear: assessing only the polish of the final product obscures the actual learning process. That polished piece could show that true learning took place, represent partial learning subsidised by system support, or suggest that the student completely bypassed cognitive engagement with the language. The most severe danger emerges here where learners might appear highly proficient based on their polished deliverables, leaving educators struggling to determine if any actual linguistic structures have been genuinely internalised.

For this reason, more robust evidentiary standards are required. Drawing on research on retention and transfer, this paper treats AI-free performance, delayed performance, and contextual transfer to novel situations as stronger indicators of actual learning (Barnett & Ceci, 2002; Roediger & Karpicke, 2006). Tests like these leave less room for confusion about what was co-produced with AI assistance and what the student independently commands.

Table 1 clarifies this distinction by comparing performance support and learning support in terms of their main orientation, evidence, roles, risks, and evaluation criteria. It shows why improvement in the final product should not, by itself, be treated as sufficient evidence of durable learner capability.

Table 1 Performance Support Versus Learning Support in Generative AI-Assisted EFL Learning

Dimension	**Performance Support**	**Learning Support**
Main orientation	Completing the task	Developing durable capability
Typical evidence	Fluent or accurate AI-assisted output	AI-free reuse, delayed performance, and transfer
Role of AI	Generator, corrector, editor, or model provider	Scaffold for noticing, explanation, evaluation, retrieval, and reconstruction
Learner role	Accepting, selecting, or submitting improved output	Analysing, explaining, revising, retrieving, and adapting language
Main risk / trade-off	Outcome ambiguity and illusion of competence	Slower visible output improvement and higher learner effort, but stronger capability evidence
Key evaluation question	Is the final product better?	Can the learner retain, reuse, or transfer the knowledge without AI later?

Note. Developed by the first author to distinguish visible task accomplishment from durable learner capability in generative AI-assisted EFL learning.

Reframing the Performance-Learning Gap Through Related Information Systems Theories

To address the performance-learning gap, this paper integrates four complementary perspectives. Theoretical perspectives were selected according to three criteria: direct relevance to the distinction between AI-assisted performance and durable learning, complementary explanatory roles, and applicability to generative AI-assisted EFL use. Effective Use addresses whether technology use leads to meaningful outcomes beyond mere system use; Task-Technology Fit helps distinguish fit for immediate task performance from fit for durable learning; socio-technical configuration explains how task design, teacher guidance, assessment, governance, and institutional conditions shape AI use; and the SECI model provides a knowledge-management perspective on internalisation. Taken together, these perspectives provide an integrated explanation of the performance-learning gap. Socio-technical configurations shape AI use patterns, which may activate or bypass the learning mechanisms discussed below. The SECI model offers a complementary perspective on knowledge internalisation, while Effective Use and Task-Technology Fit provide criteria for distinguishing immediate task success from durable learner capability. This integration connects system-level conditions with learner-level learning processes and outcomes, a relationship not fully explained by any single perspective alone.

Initially, the Theory of Effective Use provides a crucial perspective. This shifts focus away from mere system adoption to actually benefiting from its use. Burton-Jones and Grange (2013) say that effective use means leveraging a system to attain specific goals, rather than logging high frequency or even accessing it often. This distinction is acutely relevant in generative AI-assisted EFL learning. Consider a scenario where a student consults an AI source frequently and employs the software merely as a proxy author to generate texts, leaving no room for personal growth. Conversely, engaging with the AI less but critically, by comparing, explaining, revising, and then recycling what was learned, can build actual knowledge. The critical factor is whether the student internalises the linguistic knowledge to apply it later without assistance.

Furthermore, Task-Technology Fit helps explain why generative AI strongly aligns with certain performance goals but proves inadequate for learning objectives when it comes to EFL learning. Since AI can seamlessly

execute the mechanical requirements of text generation by producing polished final drafts, its performance fit might appear perfect. Yet, EFL duties aim to foster internal capability within students, not simply to produce polished deliverables. Crafting a sophisticated piece of work should inherently force learners to engage in cognitive planning, nuanced lexical selection, and error noticing. So, this model should consider both immediate task accomplishment (performance fit) and the strength of ongoing skill-building (learning fit).

Fundamentally, this nuance does not imply that the original theory is flawed for educational contexts; rather, it expands the concept. Goodhue and Thompson (1995) noted IT adds value if utilised appropriately for specific tasks. In the context of generative AI-assisted learning, how the "task" is defined dictates the adequacy of the fit. If the primary objective is rapid task execution, the fit is strong. However, if the task inherently demands the internal development of learner abilities, leveraging the technology for genuine educational gains becomes a highly complex endeavour.

Finally, turning to socio-technical configuration. Why does the same tool produce positive outcomes in one environment but fail in another? It hinges on the broader socio-technical context, the task demands, teacher guidance, school policies, and assessment criteria. For example, one course may require students to draft independently first, utilise AI for targeted feedback next, formally justify their revisions, and conclude with AI-free assessments. Another cohort might bypass these reflective stages entirely. Research on EFL students' prompt engineering similarly indicates that learners' purposes and activity contexts shape their use of generative AI, reinforcing the need for differentiated guidance and task scaffolding (Woo et al., 2025). Clearly, the first setup configures the technology as a reflective scaffold, whereas the latter positions AI as a substitute producer. This is consistent with the view that technologies are enacted differently through situated use (Baxter & Sommerville, 2011; Orlikowski, 2000).

Overall, AI-assisted learning is best understood as a whole comprising the interaction among the learner, the algorithm, the task, the pedagogy, the assessment, and the ultimate evidence of capability. Consequently, the central debate should shift away from whether AI should be banned, focusing instead on how its deployment can be systematically configured in the classroom environment.

Mechanisms, Use Patterns, and Propositions

The performance-learning gap in AI-based learning could widen or narrow depending on specific cognitive mechanisms. This study builds on second language acquisition research that focuses on noticing and the value of output for learning. The mechanisms discussed here do not constitute a new learning theory. Rather, they bring together established ideas from second language acquisition and cognitive psychology to explain how AI use might lead to lasting knowledge gains. This paper identifies five mechanism clusters: noticing and realisation, self-explanation, evaluation, retrieval, and reconstruction. These mechanisms are synthesised from research on noticing and output in second language acquisition (Schmidt, 1990; Swain, 1995; Swain & Lapkin, 1995), self-explanation (Chi et al., 1994), and retrieval practice (Roediger & Karpicke, 2006). Evaluation and reconstruction are treated in this paper as integrative mechanisms that connect these established learning processes to generative AI-assisted EFL use.

These mechanisms do not attempt to construct a grand theory covering every aspect of language learning. Instead, they serve as a practical bridge connecting how students use AI with the skills they develop. Noticing occurs when learners consciously focus on language forms, meanings, and mistakes. Realisation refers to the learner's recognition of why a noticed form, error, or mismatch matters and how their understanding or output should be revised. Realisation is therefore treated as a process that follows noticing. In an AI-assisted EFL setting, students might perceive linguistic gaps when comparing their work to AI feedback or analysing why something was changed. However, if AI silently rectifies linguistic errors and students passively accept those changes, noticing and realisation may remain unactivated.

Self-explanation means students actively articulate why a certain grammatical correction or lexical choice is structurally appropriate (Chi et al., 1994). Evaluation demands that students critically judge whether the algorithmic suggestions align with the assignment's context. These active steps matter precisely because fluent AI output is not always factually correct or contextually appropriate.

Retrieval and reconstruction tie AI use to long-term learning. Retrieval requires learners to retrieve vocabulary or grammar rules from memory without AI help. Meanwhile, reconstruction forces the student to rebuild language from memory or to use a learned structure in a fresh context. Both are vital since the goal in EFL is more than just passively recognising fluent prose on a screen; learners need to be able to use it themselves later on.

These mechanisms depend on active learner processing. Cognitive offloading becomes relevant when AI use reduces or removes the mental work needed to activate them (Risko & Gilbert, 2016). This involves outsourcing mental labour to external instruments to alleviate internal cognitive strain. In EFL classes, students might rely heavily on AI to craft paragraphs, correct grammar, or improve the apparent coherence of their prose. While offloading cognitive strain may yield rapid, high-quality deliverables, it may also bypass essential cognitive work, including generating independent answers and evaluating linguistic choices. When these processes are sidelined, durable capability formation may be severely compromised.

In this study, AI use that prioritises rapid task completion while reducing students' internal processing is identified as a "delegation-heavy" approach. This pattern does not arise solely from academic dishonesty or laziness; students may also experience linguistic insecurity, time constraints, or poorly designed assessments. Therefore, it is not merely a matter of individual behaviour but also a socio-technical issue.

Conversely, environments must cultivate learning-oriented use patterns that demand genuine cognitive engagement. Explain-first use requires an initial independent attempt before AI consultation and therefore supports retrieval and self-explanation. Compare-and-revise use focuses students on the differences between their drafts and AI suggestions and therefore supports noticing, realisation, and evaluation. Revise-with-justification use requires students to defend their accepted edits and therefore supports evaluation and reconstruction. Furthermore, retrieval-after-AI tasks require learners to reproduce or adapt language without AI and therefore support retrieval and transfer-oriented reconstruction.

Table 2 summarises the AI use patterns discussed above, the learning mechanisms they are expected to activate or weaken, and their likely effects on the performance-learning gap.

Table 2 Student AI Use Patterns, Mechanisms, and Expected Effects on the Performance-Learning Gap

Student AI use pattern	**Likely mechanism or learner process**	**Expected effect on the gap**
Delegation-heavy use	Reduced retrieval, weak evaluation, low revision ownership	Likely to widen gap
Copy-edit-submit use	Limited noticing and self-explanation	Likely to widen gap
Explain-first use	Self-explanation and learner agency	More likely to narrow gap
Compare-and-revise use	Noticing, evaluation, and revision ownership	More likely to narrow gap
Revise-with-justification use	Evaluation and reconstruction	More likely to narrow gap
Retrieval-after-AI use	Retrieval and transfer-oriented reconstruction	More likely to narrow gap

Note. Developed by the first author to show how different student AI use patterns are expected to activate or weaken learning mechanisms and thereby narrow or widen the performance–learning gap.

These learning-oriented use patterns allow AI to support the learning process while deliberately activating mechanisms associated with durable learning.

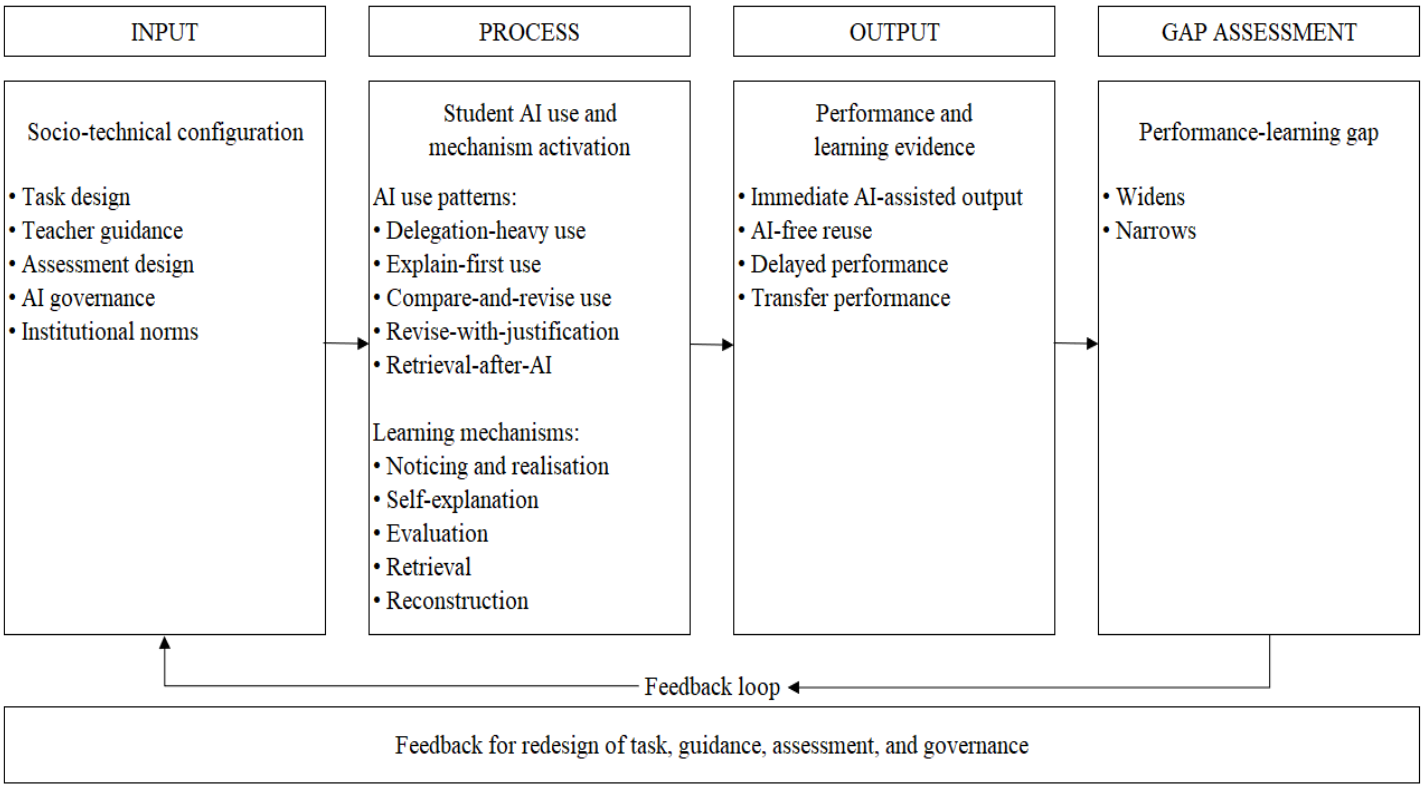

Figure 1. An Input-Process-Output-Feedback Framework of the Performance-Learning Gap in Generative AI-Assisted EFL Learning

Figure 1 integrates the proposed relationships into an input-process-output-feedback framework. The input consists of socio-technical conditions, including task design, teacher guidance, assessment, AI governance, and institutional norms. These conditions shape student AI use patterns and the activation or weakening of learning mechanisms. The output includes immediate AI-assisted performance and evidence of independent capability through AI-free reuse, delayed performance, and transfer. The divergence between these outcomes informs the gap assessment and provides feedback for redesigning instructional conditions. Future studies could examine AI use patterns through interaction records and drafts, investigate learning mechanisms through learner explanations, and compare AI-assisted performance with subsequent AI-free, delayed, and transfer performance under comparable task conditions. Based on this framework, five propositions are developed:

- P1. From an Effective Use perspective, evaluations relying solely on AI-assisted task performance are expected to create greater outcome ambiguity regarding independent learner capability than evaluations that also incorporate AI-free reuse, delayed performance, and transfer.
- P2. Consistent with the concept of cognitive offloading, delegation-heavy AI use is expected to widen the performance-learning gap when it displaces the learner processing necessary for noticing and realisation, self-explanation, evaluation, retrieval, and reconstruction.
- P3. Learning-oriented AI use patterns, including explain-first use, compare-and-revise use, revise-with-justification use, and retrieval-after-AI use, are expected to narrow the performance-learning gap to the extent that they activate noticing and realisation, self-explanation, evaluation, retrieval, and reconstruction.
- P4. From a socio-technical perspective, configurations that require independent drafting, transparent revision processes, justified AI use, and AI-free follow-up assessments are expected to promote learning-oriented AI use patterns rather than delegation-heavy use.
- P5. Extending Task-Technology Fit to learning contexts, configurations that align AI support with capability-development activities, such as learner-generated explanations, critical evaluation, and retrieval, are expected to produce stronger AI-free delayed performance and transfer than configurations optimised solely for immediate output quality.

In summary, the central issue is not whether AI is inherently good or bad for language acquisition.

Implications and Conclusion

This article enriches Information Systems scholarship by looking at the performance-learning gap as a form of outcome ambiguity in using generative AI. When AI is directly involved in creating the end product, it may obscure how much the human user has actually internalised, making the final product an unreliable proxy for independent capability.

Furthermore, this analysis pushes the boundaries of the Theory of Effective Use. If the sole institutional objective is task completion, submitting an AI-improved text technically constitutes effective use. However, if education is the primary goal, evaluating success by the polished piece of work alone is a profound shortcoming. Truly effective educational use must be measured by subsequent AI-free reuse, delayed performance, and transfer. This radically shifts the metric of success away from usage frequency or immediate output quality toward durable capability formation.

The analysis also extends Task-Technology Fit by distinguishing between performance fit and learning fit. While generative models excel at autonomously generating fluent linguistic outputs, genuine learning fit strictly depends on whether the interaction activates the mental processing necessary for memory retention and transfer. This distinction holds profound significance for IS research, since generative AI might appear highly effective if evaluation focuses solely on surface-level outcomes.

On a practical level, the primary focus should not merely be on detecting or outlawing AI use. Instead, intentionally configuring AI usage through rigorous task design, instructional guidance, and process-based accountability would be a vastly more productive paradigm. Assignments must systematically require students to attempt drafts prior to algorithmic prompting, compare AI feedback against their original text, justify accepted edits, actively reject inappropriate algorithmic suggestions, and conclude with analogue, AI-free follow-ups. Assessments must transition from grading static final projects to evaluating process evidence and independent capability. Recent work on GenAI-empowered assessment also supports the redesign of assessment activities so that students can select, use, and reflect on appropriate AI tools (Foung et al., 2024). In this paradigm, assessment acts not merely as a passive measuring stick, but as active socio-technical architecture shaping how students engage with technology, aligning perfectly with broader policy discourse regarding human-centric AI governance.

Looking ahead, more research could test this framework across multiple vectors. First, field studies should contrast environments permitting delegation-heavy use against classrooms strictly enforcing explain-first or compare-and-revise configurations to measure differing outcomes. Second, mechanism-focused studies need to conclusively determine whether noticing and realisation, self-explanation, evaluation, retrieval, and reconstruction mediate the relationship between AI interaction and durable learning. Third, longitudinal studies must isolate the retention and transferability of knowledge acquired specifically during AI-assisted learning sessions. As the present analysis focuses primarily on writing and related text-based EFL tasks, future research should also examine whether the proposed relationships hold for speaking and listening.

In conclusion, while AI can undeniably substantially improve immediate foreign language output, equating that machine-aided performance with a durable learning victory is a problematic conceptual assumption. To harness this technology for genuine capability formation, institutions must demand workflows that prevent the outsourcing of cognitive labour and stimulate internal retrieval rather than silence it. Through IS views on use, fit, and overall system designs, the ultimate challenge lies in systematically chaining temporary, digitally-mediated performance spikes to permanent, independent human cognition.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Declaration of Generative AI Use

During the preparation of this paper, the first author used the generative AI tool ChatGPT (OpenAI) for two purposes: (1) translating drafts written by the first author from Japanese into English, and (2) suggesting alternative phrasings and stylistic refinements to improve clarity, coherence, and academic tone. All prompts provided to the AI tool were based on the first author's own original ideas and text. All AI-generated translations and phrasing suggestions were critically reviewed, revised, and, where necessary, rewritten by the first author. The authors take full responsibility for the accuracy, interpretation, and originality of the content.

ACKNOWLEDGMENT

The first author would like to express his sincere gratitude to his co-supervisor, Dr. Suraya Miskon, for her valuable guidance and support throughout the research process.

REFERENCES

1. Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. Psychological Bulletin, 128(4), 612–637. https://doi.org/10.1037/0033-2909.128.4.612
2. Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. Interacting with Computers, 23(1), 4–17. https://doi.org/10.1016/j.intcom.2010.07.003
3. Burton-Jones, A., & Grange, C. (2013). From use to effective use: A representation theory perspective. Information Systems Research, 24(3), 632–658. https://doi.org/10.1287/isre.1120.0444
4. Chi, M. T. H., De Leeuw, N., Chiu, M.-H., & Lavancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18(3), 439–477. https://doi.org/10.1207/s15516709cog1803_3
5. Fan, Y., Tang, L., Le, H., Shen, K., Tan, S., Zhao, Y., Shen, Y., Li, X., & Gašević, D. (2025). Beware of metacognitive laziness: Effects of generative artificial intelligence on learning motivation, processes, and performance. British Journal of Educational Technology, 56(2), 489–530. https://doi.org/10.1111/bjet.13544
6. Foung, D., Lin, L., & Chen, J. (2024). Reinventing assessments with ChatGPT and other online tools: Opportunities for GenAI-empowered assessment practices. Computers and Education: Artificial Intelligence, 6, Article 100250. https://doi.org/10.1016/j.caeai.2024.100250
7. Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. Language Learning & Technology, 26(2), 5–24. https://doi.org/10.64152/10125/73474
8. Goodhue, D. L., & Thompson, R. L. (1995). Task-Technology fit and individual performance. MIS Quarterly, 19(2), 213–236. https://doi.org/10.2307/249689
9. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
10. Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537–550. https://doi.org/10.1177/00336882231162868
11. Koriat, A., & Bjork, R. A. (2005). Illusions of competence in monitoring one's knowledge during study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 31(2), 187–194. https://doi.org/10.1037/0278-7393.31.2.187
12. Liu, Z.-M., Hwang, G.-J., Chen, C.-Q., Chen, X.-D., & Ye, X.-D. (2026). Integrating large language models into EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. Computer Assisted Language Learning, 39(3), 466–490. https://doi.org/10.1080/09588221.2024.2389923
13. Miao, F., Holmes, W., Huang, R., & Zhang, H. (2023). Guidance for generative AI in education and research. UNESCO. https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research
14. Nonaka, I., Toyama, R., & Konno, N. (2000). SECI, ba and leadership: A unified model of dynamic knowledge creation. Long Range Planning, 33(1), 5–34. https://doi.org/10.1016/S0024-6301(99)00115-6

15. Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. Organization Science, 11(4), 404–428. https://doi.org/10.1287/orsc.11.4.404.14600
16. Polakova, P., & Ivenz, P. (2024). The impact of ChatGPT feedback on the development of EFL students' writing skills. Cogent Education, 11(1), Article 2410101. https://doi.org/10.1080/2331186X.2024.2410101
17. Risko, E. F., & Gilbert, S. J. (2016). Cognitive offloading. Trends in Cognitive Sciences, 20(9), 676–688. https://doi.org/10.1016/j.tics.2016.07.002
18. Roediger, H. L., III, & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. Psychological Science, 17(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x
19. Schmidt, R. W. (1990). The role of consciousness in second language learning. Applied Linguistics, 11(2), 129–158. https://doi.org/10.1093/applin/11.2.129
20. Swain, M. (1995). Three functions of output in second language learning. In G. Cook & B. Seidlhofer (Eds.), Principle and practice in applied linguistics: Studies in honour of H. G. Widdowson (pp. 125–144). Oxford University Press.
21. Swain, M., & Lapkin, S. (1995). Problems in output and the cognitive processes they generate: A step towards second language learning. Applied Linguistics, 16(3), 371–391. https://doi.org/10.1093/applin/16.3.371
22. Teng, M. F. (2024). "ChatGPT is the companion, not enemies": EFL learners' perceptions and experiences in using ChatGPT for feedback in writing. Computers and Education: Artificial Intelligence, 7, Article 100270. https://doi.org/10.1016/j.caeai.2024.100270
23. Woo, D. J., Guo, K., & Susanto, H. (2025). Exploring EFL students' prompt engineering in human–AI story writing: An activity theory perspective. Interactive Learning Environments, 33(1), 863–882. https://doi.org/10.1080/10494820.2024.2361381
24. Zou, S., Guo, K., Wang, J., & Liu, Y. (2026). Investigating students' uptake of teacher- and ChatGPT-generated feedback in EFL writing: A comparison study. Computer Assisted Language Learning, 39(4), 1062–1091. https://doi.org/10.1080/09588221.2024.2447279