

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

<https://llrjournal.com/index.php/11>

**Integrating a SAMR-Based Generative AI Framework to
Enhance ESL Learners' Essay Writing Skills and Critical
Thinking: A Classroom-Based Study**



¹Seemab Jamil Ghouri

²Amna Zahid

³Yaghsha Sahar

¹English Lecturer, University of Management and Technology, Sialkot, Punjab, Pakistan
seemab.jamil@skt.umt.edu.pk

²English Lecturer, GC Women University Sialkot, Punjab, Pakistan
amna.zahid@gcwus.edu.pk

³English Lecturer, Grand Asian University, Sialkot, Punjab, Pakistan.
yaghsha.sahar@gaus.edu.pk

Abstract

The use of generative artificial intelligence (AI) tools in English as a Second Language (ESL) academic writing is becoming commonplace, but their use in the classroom is not always systematic. This study investigated the perceptions of 34 undergraduate ESL students of the integration of generative AI, specifically ChatGPT, into the teaching of essay writing in four theoretically informed dimensions: AI acceptance, writing self-efficacy, critical thinking, and integration of generative AI in the context of SAMR. The composite reliabilities were acceptable to good ($\alpha = .63-.86$; 15-item overall scale $\alpha = .83$), except for a two-item acceptance subscale ($\alpha = .31$) using a descriptive-correlational survey design with a 15-item, five-point Likert questionnaire. Students' writing self-efficacy ($M = 3.91$), perceived critical-thinking engagement ($M = 3.91$), perceived SAMR-based integration ($M = 3.87$), and acceptance score ($M = 3.65$) were all high, and 76.5% planned to continue using AI for future academic writing. The SAMR-based integration was highly correlated with critical thinking ($r = .77$) and moderately-to-strongly correlated with acceptance ($r = .63$), and an independent-samples t-test showed that there was a significant difference between the self-efficacy of male and female students ($p = .003$). Interestingly, intention to continue using AI was significantly related to critical-thinking engagement ($r = .39, p = .025$), but not to acceptance alone ($r = .22, p = .22$). The results indicate that the depth of critical engagement, not just ease of use, is the strongest predictor of continued use of AI, providing empirical evidence for a staged approach to AI pedagogy, informed by SAMR, over blanket permission or prohibition policies.

Keywords: SAMR Model, generative AI, Essay Writing Skills, Critical Thinking, Writing Instruction.

Introduction

The use of generative AI tools like ChatGPT and Google Gemini has become commonplace in undergraduate academic writing, for generating ideas, drafting, paraphrasing, and revising (Anwer et al., 2025; Muslimin et al., 2024). The

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

educational impact of this transition is mixed: some studies have found improvements in language accuracy, coherence, and fluency, while others have noted potential issues with cognitive offloading and students relying on AI-generated content without critically evaluating it (Gerlich, 2025; Hong et al., 2025; Lee et al., 2025; Zakaria et al., 2025). One theme that emerges from this literature is that the impact of AI on writing quality is not simply a function of having access to the tool, but rather of how access to the tool is intentionally designed into instruction.

One framework for this integration is the SAMR model, which outlines a continuum from simple substitution of a traditional writing tool to activities that would be unimaginable without the technology (Hamilton et al., 2016). Recent research that has linked SAMR to critical thinking and technology acceptance frameworks suggests that the higher the SAMR level, the more cognitive engagement, and the more that AI is used as a scaffold to support increasingly more complex thinking, rather than as a shortcut to it (Gonsalves, 2024; Keys-Harris, 2026). In addition, there are parallel lines of research on writing self-efficacy (Bruning et al., 2002) and the Technology Acceptance Model (Davis, 1989) that indicate that perceived usefulness, ease of use, and confidence all contribute to the extent to which the use of AI leads to meaningful learning gains or superficial dependence.

Although there has been an increase in the use of SAMR-based AI pedagogy, there is limited empirical evidence that directly connects students' perceived SAMR-level engagement with their acceptance, self-efficacy, and critical-thinking behaviour, especially in ESL settings. The majority of the studies available focus on these constructs individually, not on how they are related to each other in a single cohort. This study aims to fill that gap by reporting survey data from a class of ESL undergraduates in Pakistan who had the opportunity to use generative-AI to write essays, and by testing the relationships between acceptance, self-efficacy, critical thinking, and SAMR-based integration, as well as between these factors and students' intention to continue using AI.

Objectives of the Study

The study has three objectives:

1. To examine the use of generative AI in the context of SAMR impacts the

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

writing of ESL learners.

2. To investigate the impact of SAMR-based AI-assisted instruction on ESL learners' critical thinking disposition.
3. To explore ESL learners' perceptions of AI-supported essay writing at the four SAMR levels.

Research Questions of the Study

These three objectives correspond to three research questions that organize the remainder of the study:

1. How does the integration of SAMR-based AI have on students' essay writing?
2. What impact does AI-assisted instruction have on students' critical thinking ability?
3. What is students' perception of the use of generative AI in academic writing?

Scope and Delimitations

The study is limited to 30-50 undergraduate ESL students in a single institution who are taking a Research and Academic Writing course, specifically the essay-writing unit of this course. It explores four stages of AI integration in essay writing over one semester and does not cover other language skills (such as listening or speaking) or other AI tools beyond essay writing. Findings are thus meant to inform, not to universally generalize, ESL essay-writing pedagogy.

Literature Review

The use of Generative AI in Academic Writing

There is a large and growing amount of recent research that suggests generative AI tools can enhance the surface-level and, to a lesser degree, structural components of student writing. In the context of ESL essay writing, quasi-experimental studies indicate that AI-enhanced teaching, which explicitly assigns lower-order tasks like spelling and basic grammar to the AI tool, leaves room for higher-order thinking, analysis, and evaluation (Hong et al., 2025). Trust in ChatGPT is a key factor in the perception of its writing enhancement and continued use, especially in resource-limited, ESL-heavy environments (Anwer et al., 2025). However, these benefits are consistently reported as contingent and not automatic: the same body of research highlights that gains are contingent upon the way in which AI is integrated into a

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

learning task, and not simply the availability of the tool (Muslimin et al., 2024).

Another trend in this literature is the stage of the writing process that AI is most effective at assisting with. In the current classroom, the most common areas where AI tools are used to support feedback, revision, and language polishing, whereas earlier planning and later publishing stages are relatively under-researched and under-used in the classroom. This imbalance is relevant to essay writing in particular, as the planning of an argument (what claim to make and how to support it) is arguably the most critical thinking stage, but is the least supported by structured AI use. Feedback-focused AI tools are not enough to fill the gap, as the Modification and Redefinition levels proposed in this study aim to do, by extending AI support into the planning stage.

The use of Generative AI and Critical Thinking

The application of Generative AI and Critical Thinking. AI can be used to scaffold learning, but there are many recent studies that warn of overreliance on AI that could lead to less analytical engagement and independent thinking. The survey results from knowledge workers revealed that self-reported cognitive effort decreased with the use of generative AI, with some of this decrease in effortful thinking being attributed to the confidence in the output of generative AI (Lee et al., 2025). A similar measurable relationship between the use of AI tools and a decrease in critical thinking engagement has been reported by Gerlich (2025), who found that the extent to which users rely on the AI tool without checking it independently moderates this relationship. In this article, Gonsalves (2024) revisits Bloom's Taxonomy in the context of generative AI and suggests that AI use at the lower cognitive levels of remembering and understanding is not likely to foster analysis, evaluation, or creation, while AI use intentionally aimed at provoking critique, comparison, and synthesis can support, but not replace, higher-order thinking. This is at the heart of the current study's argument for the use of SAMR as a sequencing tool: not all uses of AI are equally impactful for critical thinking, and the SAMR levels provide a principled means of distinguishing shallow from deep engagement.

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

The use of Bloom's Taxonomy and Cognitive Demand in AI-Integrated Writing

Recent research has combined SAMR with the revised taxonomy of Bloom to make the cognitive demand a task places on a learner explicit (Gonsalves, 2024; Keys-Harris, 2026). In this pairing, Substitution and Augmentation are typically linked to lower levels of remembering, understanding, and applying, as the task of writing itself does not change significantly when AI takes over a portion of the task. Modification and Redefinition, on the other hand, are linked to analysing, evaluating, and creating, as these stages involve questioning AI-generated ideas, considering alternative viewpoints, and generating a new product that demonstrates independent thinking and not simply the acceptance of AI-generated ideas. This mapping serves as the theoretical foundation for this study's hypothesis that gains in writing and critical thinking will not be consistent across SAMR stages, but will follow the cognitive demands of each stage.

Writing Self-Efficacy in AI-Assisted Contexts

Following the basic formulation of Bruning et al. (2002), writing self-efficacy, that is, learners' confidence in their ability to plan, draft, organise and revise a piece of writing, is one of the more consistent predictors of writing performance in the composition literature. This construct seems to be shifting rather than going away: students' confidence is shifting from their ability to compose to their ability to guide, ask questions, and edit AI-generated suggestions. The change is significant for instructional design, as an intervention that boosts essay scores but does not also increase students' confidence in their own writing could yield short-term results that are not sustainable without the assistance of AI.

Technology Acceptance and AI Adoption in ESL Writing

The Technology Acceptance Model (TAM) (Davis, 1989) remains the most widely used framework to understand student uptake of AI writing tools, which is usually measured by perceived usefulness, perceived ease of use, and behavioural intention. Recent studies in the Pakistani context of higher education have focused specifically on the role of trust in the TAM, finding that trust is a key predictor on top of the traditional TAM variables: students who trust the output of ChatGPT are more likely to report actual improvement in their writing and to express their intention to continue

using the tool for future academic tasks (Anwer et al., 2025). Another large sample study of university students in Pakistan also revealed that perceived usefulness and associated TAM constructs are significant predictors of the adoption of generative AI in higher education (Shahzad et al., 2024). In addition, cross-national comparative studies with Pakistani and Saudi samples indicate that the strength of the TAM constructs' relationship with actual engagement is moderated by cultural context (Alharbi et al., 2025).

The SAMR and SAMR-AI Integration in Language Education

Although the SAMR model was not originally created for the purpose of integrating AI, it has recently gained popularity as a framework for understanding how AI can be used to enhance teaching and learning. The SAMR model's four levels have proven to be a useful framework for structuring the impact of AI tools on task design in various national contexts of language instruction, as shown in recent applications to business language instruction (Herget & da Silva Ferreira, 2025). A 2026 EDUCAUSE framework also explicitly connects SAMR to Bloom's Taxonomy, allowing teachers to identify the cognitive level of an assignment and intentionally determine how far up the SAMR continuum they want to allow AI to take an assignment (Keys-Harris, 2026). In the context of EFL writing, the results of the evaluation of the use of AI tools at each level of SAMR show that the achievement and perception of students vary significantly between the levels, with higher-order levels showing better achievement and perception (Muslimin et al., 2024).

For the purposes of this study, the four levels of SAMR are translated into four different classroom activities and AI roles in Table 1.

SAMR Level	AI Role	Representative Classroom Activity
Substitution	Direct replacement of manual proofreading	Students draft an essay conventionally; AI is used only to correct grammar and spelling.
Augmentation	Functional improvement of clarity	Students revise sentence-level wording and paragraph coherence using AI

	and structure	suggestions.
Modification	Task redesign for argument development	Students use AI-generated counterarguments to test and revise their thesis statements.
Redefinition	Enables tasks otherwise not possible	Students engage AI in structured critical debate and iteratively redesign the full essay.

Methods

Research Design

The design of this study was descriptive-correlational quantitative survey. The design is appropriate for the study's purpose of quantifying students' perceptions across four dimensions and testing the associations among them, but not for making causal claims about the impact of AI on writing performance.

Participants

The subjects were 34 undergraduate students of BS English who had written essays with the help of generative AI tools in their academic studies. All 34 students who were invited completed the questionnaire. Participant characteristics are summarized in Table 1.

Table 1. Participant Characteristics (N = 34)

Characteristic	Category	N	%
Gender	Female	24	70.6
	Male	10	29.4
Semester	4th	26	76.5
	5th	5	14.7
	6th	1	2.9
	Not reported	2	5.9
AI tool used most	ChatGPT	23	67.6

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

	Google Gemini	5	14.7
	Other tools	5	14.7
	Not reported	1	2.9
Intend to keep using AI	Yes	26	76.5
	No	8	23.5

Instrument

A self-developed four dimension Likert questionnaire with 15 items rated on a 5-point scale (1 = Strongly Disagree to 5 = Strongly Agree), one binary item (intention to continue using AI) and two open-ended questions were used to collect data. The four dimensions, based on the Technology Acceptance Model, writing self-efficacy theory, and critical-thinking disposition theory, and organized around the SAMR model, were: (a) AI Acceptance (2 items: ease of use, perceived efficiency), (b) Writing Self-Efficacy (5 items: confidence, idea organization, revision, and expression), (c) Critical Thinking (4 items: evaluating, comparing, and selectively adopting AI suggestions), and (d) SAMR-Based Integration (4 items: AI's role in correcting language, improving clarity, strengthening arguments, and improving overall writing, corresponding to the Substitution/Augmentation, Augmentation, Modification, and Redefinition levels of SAMR).

Procedure

After engaging in essay writing activities with the assistance of AI, students were given class time to fill out an anonymous Google Forms survey. Students were informed that their responses would be used only for research purposes and participation was voluntary.

Data Analysis

Likert responses were scored on a numerical scale (1-5). Internal consistency was evaluated using Cronbach's alpha for each dimension and for the total 15 item scale with listwise deletion for missing values. Descriptive statistics (M, SD, minimum, maximum) were calculated at the item and composite level. Pearson correlations were used to explore the relationships between the four composite dimensions and between each dimension and intention to continue using AI (point-biserial correlation, since

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

this was binary coded). Independent-samples t-tests (Welch's correction) were used to compare composite scores between gender and between AI tool used (ChatGPT vs. other tools). The qualitative results were analyzed and presented in the form of descriptive themes to provide context to the quantitative results.

Results

The internal consistency of the Writing Self-Efficacy was good ($\alpha = .857$), while SAMR-Based Integration and Critical Thinking were acceptable ($\alpha = .694$ and $.629$, respectively). The overall scale (15 items) had good reliability ($\alpha = .825$). The two-item Acceptance subscale had poor reliability ($\alpha = .310$); this is discussed as a limitation below and results for this dimension should be interpreted with caution.

Table 2. Internal Consistency by Dimension

Dimension	Items	Cronbach's α	N
AI Acceptance	2	.310	34
Writing Self-Efficacy	5	.857	33
Critical Thinking	4	.629	33
SAMR-Based Integration	4	.694	33
Overall scale	15	.825	32

Descriptive Statistics by Dimension

Table 3 shows composite descriptive statistics for the four dimensions. Writing Self-Efficacy and SAMR-Based Integration had the highest mean scores ($M = 3.91$ for both), followed closely by SAMR-Based Writing ($M = 3.89$). AI Acceptance had a relatively low, but positive, mean ($M = 3.65$) and the largest relative range when the low reliability is considered.

Table 3. Descriptive Statistics by Dimension

Dimension	M	SD
AI Acceptance	3.65	0.54
Writing Self-Efficacy	3.91	0.63
Critical Thinking	3.91	0.55

SAMR-Based Integration	3.87	0.56
------------------------	------	------

The item-level statistics are presented in Table 4. In Critical Thinking, students most strongly agreed that AI encouraged them to consider different perspectives while writing ($M = 4.09$, $SD = 0.62$) and that they compared AI-generated ideas with their own ($M = 4.00$, $SD = 0.98$). For Writing Self-Efficacy, the means for confidence in writing an academic essay and expressing ideas clearly were both $M = 3.97$. In SAMR-Based Integration, AI's impact on grammar correction and on enhancing clarity was both equally high ($M = 3.94$), and the lowest score was for its contribution to overall writing improvement ($M = 3.74$) - as noted earlier, the highest SAMR levels are the most difficult to implement in practice (De Castro & Guia, 2025).

Table 4. Item-Level Descriptive Statistics

Item	Dimension	M	SD	Min	Max
AI was easy to use during essay writing	Acceptance	3.74	0.62	3	5
AI made the essay writing process more efficient	Acceptance	3.56	0.79	1	4
I feel confident writing an academic essay in English	Self-Efficacy	3.97	0.83	2	5
I can organize my ideas clearly when writing an essay	Self-Efficacy	3.91	0.87	2	5
I can revise my essay effectively after receiving feedback	Self-Efficacy	3.85	0.70	3	5
I can revise my essay effectively after receiving feedback (2)	Self-Efficacy	3.82	0.77	2	5
I can express my ideas clearly in academic writing	Self-Efficacy	3.97	0.76	2	5
I evaluated AI-generated suggestions before including them	Critical Thinking	3.67	0.82	2	5
I compared AI-generated ideas with my own before deciding	Critical Thinking	4.00	0.98	1	5

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

AI encouraged me to consider different perspectives	Critical Thinking	4.09	0.62	3	5
I selected AI suggestions only when they improved my writing	Critical Thinking	3.85	0.70	2	5
AI effectively helped correct grammar and spelling errors	SAMR Integration	3.94	0.81	2	5
AI improved the clarity and organization of my essay	SAMR Integration	3.94	0.70	3	5
AI supported me in developing stronger arguments and ideas	SAMR Integration	3.85	0.78	2	5
AI use improved my overall academic writing skills	SAMR Integration	3.74	0.75	2	5

Relationships Among Dimensions

Pearson correlations between the four composite dimensions are shown in Table 5. SAMR Based Integration was significantly related to Critical Thinking ($r = .766$) and to Acceptance ($r = .634$) and moderately related to Self-Efficacy ($r = .317$). There was a moderate correlation between Acceptance and Critical Thinking ($r = .507$). Self-Efficacy had relatively low correlations with Critical Thinking ($r = .244$) and Acceptance ($r = .347$), indicating that students' confidence in their own writing ability was relatively independent of their critical thinking and acceptance of AI.

Table 5. The Pearson Correlations among the Composite Dimensions are shown below

Dimension	Acceptance	Self-Efficacy	Critical Thinking	SAMR Integration
Acceptance	-			
Self-Efficacy	.347	-		
Critical Thinking	.507	.244	-	
SAMR Integration	.634	.317	.766	-

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

Only Critical Thinking ($r = .385$, $p = .025$) was significantly correlated with intention to continue using AI in future academic writing (coded 1 = Yes, 0 = No). Correlations with Acceptance ($r = .217$, $p = .218$), Self-Efficacy ($r = .187$, $p = .289$), and SAMR Integration ($r = .281$, $p = .107$) were not significant, but were close to significance for SAMR Integration.

Group Comparisons

Independent-samples t-tests (Table 6) showed that there was a significant difference between the male students ($M = 4.39$, $SD = 0.50$) and the female students ($M = 3.71$, $SD = 0.57$) in Writing Self-Efficacy, $t = -3.46$, $p = .003$. There was no significant difference between the genders in any other dimension. There were no statistically significant differences between the students who used ChatGPT most often ($n = 23$) and those who used other tools ($n = 11$, who used Gemini and other tools) on any dimension, indicating that perceptions were not strongly tool-specific.

Table 6. Group Comparisons by Gender and by AI Tool Used

Dimension	Female M (SD)	Male M (SD)	t	p	ChatGPT M	Other tools M	t	P
Acceptance	3.62 (0.56)	3.70 (0.54)	-0.37	.718	3.59	3.75	-0.79	.440
Self-Efficacy	3.71 (0.57)	4.39 (0.50)	-3.46	.003*	3.77	4.15	-1.71	.104
Critical Thinking	3.90 (0.43)	3.94 (0.81)	-0.17	.868	3.90	3.85	0.20	.848
SAMR Integration	3.81 (0.50)	4.02 (0.68)	-0.85	.408	3.82	3.92	-0.52	.607

A total of 34 students responded to each of the open-ended questions, with 28 of them giving a substantive response. Responses were analysed line by line and inductively categorised into themes, with some responses addressing more than one theme, so percentages do not add up to 100%. Themes identified for the most helpful aspect question are summarized in Table 7 and the challenges/concerns question in Table 8.

Table 7. Most Helpful Aspect of Generative AI

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

Theme	n	%	Illustrative excerpt
Idea generation and structuring	17	60.7	"It helps you beat writer's block by quickly turning a blank page into a clear outline."
Grammar and language accuracy	11	39.3	"Chatgpt helped me organize my ideas and improve my grammar which made essay writing easier and faster."
Efficiency and time-saving	9	32.1	"It helped generate ideas, improve grammar, and save time."
Vocabulary enhancement	3	10.7	"It improves my vocabulary and grammar."
Immediate, targeted feedback	3	10.7	"It provides me step by step directions and quickly checks my errors."
Access to broad information	2	7.1	"AI is helpful because it gives us required information all over the world."
Confidence / reduced stress	1	3.6	"...made the writing process faster and less stressful."

By far the most common theme was the use of AI to create and organize ideas, guiding students from a blank page to a structured outline or 'road map' for their essay. This was also referenced approximately twice as much as grammar support, the second most common theme, indicating that students used AI more as a planning and organizational aid than as a proofreading tool. A small but significant minority (one respondent, elsewhere in the same survey) qualified their use of AI as 'deliberately limited', saying they did not want to rely on AI because it 'avoids your brainstorming' - a self-regulated caution that was not prompted, but was expressed by a small minority of respondents, and that anticipates the dependency concerns raised more explicitly in the challenges data below.

Table 8. Major Challenges Using Generative AI

Theme	n	%	Illustrative excerpt
Accuracy/reliability of AI content	14	50.0	"Sometimes the information was too general or inaccurate, so I had to verify it from reliable

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

			sources."
Prompt-dependency and misread intent	6	21.4	"It does not understand my actual problem. It gives me general suggestions."
Over-reliance / dependency risk	4	14.3	"...I also think using AI too much can make students depend on it."
Originality, voice, and AI-detection risk	4	14.3	"The biggest worry is accidental plagiarism or AI detection..."
Tone/style (robotic, generic)	2	7.1	"...occasionally used words that sounded too formal or robotic..."
Cost/access barriers	2	7.1	"Unpaid version... does not provide proper direction."
Lacks human nuance / argumentative bias	2	7.1	"AI provides a good quantitative argument but can't provide a bias opinion... which is the core of essay writing."

Half of all respondents (50%) said the accuracy and reliability of AI-generated content was the primary concern, with outputs being 'too general,' sometimes factually incorrect, or needing to be manually checked. This theme, along with another secondary theme of prompt-dependency (AI 'not understanding' what the student meant, or needing a well-formed prompt to be useful), suggests a specific and correctable source of friction: students are not simply relying on AI, but are being asked to do additional verification and reformulation work that a more structured, higher-SAMR-level task design might be able to minimize. One answer also identified a sycophancy-like pattern, where the AI was agreeing with the student's position, irrespective of whether it was correct or not, which is a separate trust issue from the factual accuracy and should be tracked explicitly in future instrument design. When read together, these themes triangulate very closely with the quantitative Critical Thinking results. The same students who rated 'I evaluated AI-generated suggestions before including them' and 'I compared AI-generated ideas with my own ideas' highly ($M = 3.67$ and $M = 4.00$, respectively) independently, and without

prompting, described exactly this evaluative behaviour in their own words - checking facts, rewriting robotic-sounding phrasing, and verifying information from other sources. The qualitative data thus corroborate the quantitative finding in the Discussion below: students' overall positive attitudes towards AI are not necessarily accompanied by uncritical acceptance, but rather by a conscious and monitored attitude towards the output of AI.

Discussion

In general, students' perceptions were positive in all four dimensions, with the highest and most consistent scores in Writing Self-Efficacy and Critical Thinking. This is similar to previous studies that found that AI tools can enhance learners' confidence and engagement when used in a structured writing task, as opposed to being used freely by the students (Anwer et al., 2025; Muslimin et al., 2024). Concurrently, several findings challenge a straightforward 'students like AI' story and instead lend credence to the more nuanced, SAMR-based story that this study aimed to validate.

The highest correlation in the data set was between SAMR Based Integration and Critical Thinking ($r = .766$). Students who experienced AI at a more substitutive level also reported less evaluation, comparison, and selective adoption of AI suggestions than students who experienced AI at a more strengthening level, which involved using AI to enhance arguments and overall writing, rather than just to correct grammar. This is exactly what SAMR-based descriptions of AI pedagogy suggest: that the higher the SAMR level, the more cognitively demanding the task (Gonsalves, 2024; Keys-Harris, 2026). The relatively high correlation between SAMR-Based Integration and Acceptance ($r = .634$) indicates that perceived usefulness, rather than ease of use, is what maintains positive attitudes toward the tool, an extension of the Technology Acceptance Model's focus on perceived usefulness (Davis, 1989) to an AI-writing context.

The results were somewhat surprising as Writing Self-Efficacy was only moderately related to Critical Thinking ($r = .244$) and SAMR-Based Integration ($r = .317$). In other words, confidence in their own writing ability was not significantly related to the critical engagement with AI output. A possible interpretation is that self-efficacy in this sample is not specific to directing and evaluating AI support, but rather general

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

academic writing confidence, which is developed during a semester of academic writing. In recent extensions of self-efficacy theory to human-AI collaborative writing, one possible interpretation is that self-efficacy in this sample is not specific to directing and evaluating AI support, but rather general academic writing confidence, which is developed during a semester of academic writing. If self-efficacy and critical engagement are indeed separable, then this has a practical implication: teaching students to be confident in using AI to write does not necessarily lead to teaching them to engage critically with the writing process; they may need to be taught separately.

The connection between intention to continue using AI and the four dimensions is another refinement, if not more important. There was a significant correlation between Intention and Critical Thinking ($r = .385$, $p = .025$), but not between Intention and raw Acceptance ($r = .217$, $p = .218$). If ease of use or convenience were the primary drivers of sustained adoption, as would be expected by classic technology-acceptance models, Acceptance should have been the better predictor. Rather, students who indicated that they assessed and critically considered AI suggestions were more likely to indicate that they would continue to use the tool. This indicates that, at least in this group, the quality of engagement, which SAMR-based integration is supposed to foster, is what keeps people using SAMR, rather than convenience. This nuance builds on TAM-based studies of AI use in ESL settings (Anwer et al., 2025; Shahzad et al., 2024), which focus on acceptance as a key factor in sustaining AI use, by highlighting the importance of critical engagement.

The lack of significant differences between the ChatGPT users and the users of other tools in all four dimensions indicates that these perceptions are not strongly associated with a particular AI product. This aligns with the larger argument that the pedagogical framework in which AI is used, not the tool itself, determines learning-relevant outcomes (Herget & da Silva Ferreira, 2025) and that SAMR-based integration is not a tool-specific instructional approach, but rather a tool-agnostic approach.

The Writing Self-Efficacy (Males higher, $p = .003$) is significant, but should be interpreted with caution because of the small and unequal size of the subgroups (10 male, 24 female respondents). This is not a strong or generalizable gender effect, but

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

rather a signal to be revisited in a larger, more balanced sample; it may be due to differences in prior writing experience, differences in self-report tendencies around confidence, or cohort-specific factors not measured by this survey. The gender difference was not observed in Critical Thinking or SAMR-Based Integration, suggesting that the gender difference, if it exists, is not related to critical engagement with AI, but rather to self-reported confidence.

Lastly, the thematic analysis of open-ended responses provides insights into the reasons behind the relatively high scores on Critical Thinking items, even with the overall positive perceptions of AI. The single largest challenge theme - accuracy and reliability of AI content - and the prompt-dependency theme (AI 'not understanding' the student's intent, or one respondent's observation that the tool tends to simply agree with whatever the student proposes) likely led to the evaluative behaviours captured in the Critical Thinking items: comparing AI output with one's own ideas, and selecting suggestions only when judged to improve the writing. Students' positive attitudes toward AI do not seem to be passive, but rather a conscious and monitored approach to the results of AI, and in some cases, a self-imposed restriction on the use of AI to preserve their own brainstorming, which is the desired effect of a SAMR-informed, critical-thinking approach to AI integration.

Limitations

The study contains some limitations. First, the sample ($N = 34$) was limited to a single course, which reduces generalizability. Second, the two-item Acceptance subscale had a low internal consistency ($\alpha = .310$) and should be expanded and re-validated in future studies. Third, the cross-sectional, correlational design does not allow for causal or temporal relationships between the four dimensions to be determined; that is, it is not possible to conclude that higher perceived SAMR-level integration leads to deeper critical thinking, or that students who already think more critically perceive and use AI at a higher SAMR level.

Conclusion

The results of this study revealed that ESL undergraduate students had positive perceptions of the use of generative AI in essay writing in terms of acceptance, self-efficacy, critical thinking, and SAMR-based integration, with the highest scores being

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

for writing self-efficacy and critical thinking. The results of this study suggest that SAMR integration is linked to critical thinking, and that critical engagement, not acceptance, was a predictor of intention to continue using AI, which aligns with the idea that AI integration is a pedagogical design problem, not a binary choice of permitting or prohibiting its use. In practice, these findings indicate that teachers and institutions should plan the use of AI in a deliberate manner, starting with grammar-level support and moving towards tasks that involve students in evaluating, comparing, and critically redesigning AI-generated content. They also suggest that simply increasing students' confidence in using AI does not automatically lead to the development of the critical evaluation skills that are necessary for meaningful use.

References

- Anwer, B., Brohi, R., & Alvi, A. (2025). Trust in ChatGPT and perceived academic writing improvement: A TAM-based quantitative study in a Pakistani ESL context. *Futurity Education*, 5(4), 4-20.
- Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2002). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology*, 94(4), 668-679.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- De Castro, C. C., & Guia, A. (2025). Effectiveness of the SAMR Model for Enhancing Literature Instruction in Junior High School. *Journal of Interdisciplinary Perspectives*, 3(6). <https://doi.org/10.69569/jip.2025.211>
- Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction*. The California Academic Press.
- Gerlich, M. (2025). AI tools in society: Impacts on cognitive offloading and the future of critical thinking. *Societies*, 15(1), 6. <https://doi.org/10.3390/soc15010006>
- Gonsalves, C. (2024). Generative AI's impact on critical thinking: Revisiting Bloom's taxonomy. *Journal of Marketing Education*. <https://doi.org/10.1177/02734753241305980>
- Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and

Liberal Journal of Language & Literature Review

Print ISSN: 3006-5887

Online ISSN: 3006-5895

suggestions for its use. *TechTrends*, 60(5), 433-441. <https://doi.org/10.1007/s11528-016-0091-y>

Herget, K., & da Silva Ferreira, K. (2025). Applying the SAMR model to AI-enhanced business language instruction: Comparative insights from German and Spanish contexts. *Journal of Language and Education*, 11(3), 160-167. <https://doi.org/10.17323/jle.2025.27419>

Hong, H., Vate-U-Lan, P., & Viriyavejakul, C. (2025). Cognitive offload instruction with generative AI: A quasi-experimental study on critical thinking gains in English writing. *Forum for Linguistic Studies*, 7(7), 325-334. <https://doi.org/10.30564/fls.v7i7.10072>

Keys-Harris, A. (2026). A practical SAMR + AI framework for instructional design. *EDUCAUSE Review*.

Lee, H.-P., Sarkar, A., Tankelevitch, L., Drosos, I., Rintel, S., Banks, R., & Wilson, N. (2025). The impact of generative AI on critical thinking: Self-reported reductions in cognitive effort and confidence effects from a survey of knowledge workers. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*, Article 1121, 1-22. <https://doi.org/10.1145/3706598.3713778>

Muslimin, A. I., Mukminatien, N., & Ivone, F. M. (2024). Evaluating Cami AI across SAMR stages: Students' achievement and perceptions in EFL writing instruction. *Online Learning*, 28(2), 1-19. <https://doi.org/10.24059/olj.v28i2.4246>

Shahzad, M. F., Xu, S., & Asif, M. (2024). Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model. *British Educational Research Journal*, 51(2), 489-513. <https://doi.org/10.1002/berj.4084>

Zakaria, N. Y. K., Hashim, H., & Jamaludin, K. J. (2025). Exploring the impact of AI on critical thinking development in ESL: A systematic literature review. *Arab World English Journal*, Special Issue on Artificial Intelligence, 330-347.