Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts January-June 2026 7(1), 668-680

DIGITAL VISUALITIES AND MEDIA PRACTICES: EXPLORING AI-GENERATED NARRATIVES FOR LEARNER ENGAGEMENT IN ESL CLASSROOMS

Dr. P. Madhan [1], Dr. Julius Irudayasamy [2], Dr. M. Natarajan [3]

[1] Professor and Head, Department of English and Foreign Languages, Alagappa University, Karaikudi, Tamil Nadu, India
[2] Assistant Professor, Department of English Language and Literature, Dhofar University, Salalah, Oman
[3] Assistant Professor, Department of English and Foreign Languages, Alagappa University, Karaikudi, Tamil Nadu, India

Received 02 March 2026
Accepted 10 May 2026
Published 22 May 2026

Corresponding Author
Dr. P. Madhan,
ponmadhanrames@gmail.com

DOI
10.29121/shodhkosh.v7.i1.2026.8365

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Background: The development of visual storytelling AI generators (DALL-E, Midjourney) has become a common feature of contemporary English as a Second Language (ESL) classes. Nevertheless, the current status and evidence on potential impacts of AI-based visual stories on learners' engagement exhibit low proficiency.
Purpose: The present research aims at studying the effect of AI-based visual stories on behavioural, emotional, and cognitive engagement of ESL learners in the context of narrative writing within university settings of selected undergraduate students at random in India.
Methodology: A quasi-experiment with a mixed-methods approach was conducted among 48 B2-level undergraduate ESL learners for eight weeks using the author's AI visual storytelling tool VisNAR. Data collection methods included validated questionnaires measuring engagement both before and after the intervention, cognitive load measurement scales, task completion time records, and semi-structured interviews conducted with 12 participants. These participants voluntarily agreed to take part in the study.
Results: There was a statistically significant rise in behavioural engagement (from $M = 3.12$, $SD = 0.68$ pre to $M = 4.01$, $SD = 0.72$ post; $t(47) = 6.21$, $p < .001$, $d = 0.89$) as well as emotional engagement (from pre $M = 2.95$, $SD = 0.81$ to post $M = 3.88$, $SD = 0.79$; $t(47) = 5.94$, $p < .001$, $d = 0.78$). However, no significant changes in cognitive engagement were observed (from pre-$M = 3.45$, $SD = 0.70$ to post $M = 3.52$, $SD = 0.84$, $p = .557$). Perceived cognitive load dropped from $M = 5.2$ to $M = 4.1$ ($p = .03$). According to the results of thematic analysis of qualitative data, visual stimuli were a "spark" to initiate writing activities in 11 out of 12 cases, while cognitive overload occurred in six out of 12 cases owing to the number of visual-stimuli details; moreover, visual fixation and passive imitation of AI-generated content were reported in five out of 12 cases.
Conclusion: AI-generated visual content enhances engagement on behavioural and emotional levels, although it does not affect cognitive engagement and can replace language skills and processes with passive description.

Keywords: AI-Generated Visuals, ESL Engagement, Cognitive Load, Visual Narratives, Multimodal Literacy

1. INTRODUCTION
1.1. BACKGROUND OF THE STUDY

Generative AI technology has radically changed how digital contents are made, allowing learners at various undergraduate students, to generate advanced imagery without any special artistic skills. AI applications such as DALL-E, Midjourney, and Stable Diffusion can generate top-quality and contextualized images based on text descriptions in seconds. With this innovation, there are many implications in terms of English as a second language (ESL) classes, which were traditionally centred around text-based instructions and assignments, valuing written language above others.

There is one potential way in which AI technologies may contribute to ESL teaching and learning through visual storytelling. These types of materials represent multiple images that together form a story and thus provide students with additional support when constructing a narrative. Although standalone images and even picture sequences were already proven beneficial in terms of vocabulary acquisition and inferencing (Liu, 2016; Tomlinson, 2013), AI-generated visual narratives represent a completely new concept because they include the following elements that were not characteristic of earlier types of materials: they are created by prompting learners to provide prompts, they are iterative, and they contain certain errors or hallucinations. At the moment, however, there is very little empirical evidence regarding the impact of AI-generated visual narratives on ESL learning. The current project aims to address this gap and determine how visual narratives produced using AI affect learner engagement on three levels in accordance with the following research questions:

1.2. STATEMENT OF THE PROBLEM

Generative AI has captured widespread societal and institutional attention, especially in educational sectors, but in reality, research has not caught up-least of all in ESL classrooms. The problem lies in the incorporation of AI-generated visual narratives into classroom practices and the degree of empirical evaluation of their effects on students' engagement. There is not much research that addresses authentic engagement, covering behavioural, affective, and cognitive aspects. This is important since, according to Philp and Duchesne (2016), engagement mainly determines the efficiency of instructional design. However, without reliable evidence, teachers are unable to ensure to what extend AI-generated visuals have impacts on students: whether the former enhances the writing proficiency of the latter,

Furthermore, current academic literature mostly discusses static images, such as stock photos, illustrations, or sketches made on a blackboard. These elements can be useful; however, it is unclear whether they are applied to AI-generated visuals. The reasons include: First, AI images are different because students drive them. Students are not merely looking at a picture. They are writing prompts, tweaking them, figuring out what works. Such interventions may fundamentally alter student engagement dynamics

Secondly, AI-generated pictures can easily be edited instantaneously through generating a new picture with the introduction of a new prompt.

Thirdly, and most importantly, AI generates occasional mistakes: either an anomaly or an element which appears to be misplaced within the context. In this case, regular images will rarely display any such anomalies. This may cause confusion to students; but then again, in certain cases, it can stimulate unexpected cognitive activities.

As seen in light of these observations, several fundamental questions arise. These are certainly no minor questions.

The first question to ask is: are students working with the help of AI more actively in general? Are they spending more time in writing, using such images, or conversely foster heightened levels of learner engagement?

Next comes a more fundamental question regarding the cognitive activities. Will the inclusion of the visual element increase students' focus, or rather will it distract from the main goal of study?

The third question to ask here would be: what cannot be quantified? Perhaps, the student will merely copy-paste the picture, run into obstacles, or even think differently than before. And yet, such aspects will not be captured in statistics.

Answering these three questions can greatly contribute to addressing the deficiency in the literature.

1.3. PURPOSE OF THE STUDY

What are the specific goals of the study? First and foremost, three main goals that are related to three research questions.

In other words, the researchers wanted to find out whether students' engagement is boosted significantly through an eight-week period when they write using the AI-generated images. To verify this hypothesis, we considered the following aspects:

1) Time on task: How much time do learners spend on writing sessions? The duration was measured using the Learning Management System (LMS).

2) Tools usage: The researchers were interested in what degree of independence learners demonstrate when working with AI tools, including the ability to generate prompts, manipulate images, and try different options.
3) Self-report: Finally, the researchers conducted a valid survey asking the respondents about their level of engagement on a five-point Likert scale.

1.4. RESEARCH QUESTIONS

RQ1: To what extent do AI-generated visual narratives influence behavioural engagement (time-on-task, tool use, self-report)?

RQ2: How do AI visuals affect cognitive engagement and perceived cognitive load?

RQ3: What qualitative patterns emerge from student interaction with AI visuals?

Which cognitive processes are involved?

This question is more challenging than the previous one because it is aimed at understanding how AI-generated images affect cognitive processes, including self-regulated learning and strategy usage. In addition, we wanted to find out if the use of extra visual information may increase extraneous cognitive load or distract learners. We used the Paas Scale. It tracks changes in how much mental effort students feel. We looked for lighter cognitive "lift" in the right areas. But we also flagged any dips in the mental effort that actually helps learning.

What do students experience? As numerical data do not give much insight, our third research question concerns a more profound and complex analysis. Interviews of purposefully chosen twelve participants lasted for about twenty-five minutes and included eight questions per student, with answers transcribed verbatim. Moreover, we analyzed two hundred forty AI visuals alongside with students' written thoughts about them.

What did we observe? Is there anything particular about students' engagement with visuals? Do they feel too much pressure or inability to progress? Does it affect students’ oral presentation depending on the presence of AI visuals? The fact is that numerical data miss all these aspects. On the contrary, interviews, artefacts, and narratives reveal the qualitative context behind numbers, helping us understand both whether AI visuals affect students and why.

What does the general conclusion mean? As a result of analyzing quantitative and qualitative data within three research objectives, we get the overall analysis of the use of AI visuals as a pedagogical tool in teaching ESL writing. The main idea is that such mixed-methods analysis can be used to address the problem identified in previous studies.

2. LITERATURE REVIEW

2.1. THEORETICAL FRAMEWORK

The project draws upon the ideas of three theories to examine how the use of AI-generated visual narratives would foster the engagement of ESL learners.

Theory 1: Multimodal Learning Theory (Mayer, 2020)

Mayer's Multimodal Learning Theory will be referenced first. According to Mayer (2020), there are two ways of acquiring knowledge: through words and images. Simultaneously, the brain is able to perceive both at the same time until reaching capacity. Thus, learning is effective when both types of stimuli are processed together and help to create a cohesive representation in the learner's mind. Mayer's experimental work provides a number of guidelines in the matter: synchronize images and text, strive for simplicity and highlight the significant. Applying these recommendations to AI-generated imagery seems quite relevant because of the effect of strong image-text matching improving understanding of content. At the same time, complexity and randomness in visuals might prevent effective learning. With students controlling the content, the accuracy of image representation in terms of the corresponding text grows

Theory 2: Cognitive Load Theory (Sweller, 2011)

Cognitive Load Theory provides a framework for understanding the mental effort demands of learning tasks. The theory distinguishes three types of cognitive load:

According to Cognitive Load Theory (Sweller, 2011), three types of cognitive load must be considered when integrating AI-generated visuals into ESL writing instruction. The first type is the intrinsic cognitive load, which refers to the difficulty of the content itself; the intrinsic load of ESL narrative writing exercises is rather high because they require the use of appropriate vocabulary, grammar, and discourse structures at the same time. The second type, extraneous cognitive load, stems from inefficient instructional design; AI-generated visuals containing irrelevant or hallucinated information can become extraneous loads on learners' minds. Third, germane load represents the mental effort devoted to schema construction and deep processing that leads to meaningful learning. This brings forth an interesting dilemma. The AI-generated narrative could become too detailed and well-presented to the point where it actually inhibits the requirement for cognitive filling in of blanks. For instance, the effort put into constructing sentences and choosing the right vocabulary may become unnecessary because students will only need to explain what they see. This will lead to passivity in learners and lack of engagement that offers few opportunities for educational gains. On the other hand, this approach may be simpler for students. But that is exactly the problem. Because second language acquisition needs deep processing.

The critical question for this study is whether AI-generated visuals reduce extraneous load (by offloading mental visualization of settings and characters) without simultaneously reducing germane load (by providing such complete visuals that learners no longer need to engage in deep linguistic processing). This tension is directly tested in the present study.

Theory 3: Self-Determination Theory (Deci & Ryan, 2000)

Second, SDT provides insight into what could motivate students to continue working on the project. SDT highlights three main needs of learners: competence, autonomy and relatedness. By creating visuals and discussing them in class, the learners might fulfil their needs for efficacy, control and social connectedness to their peers. Another scenario worth examining is that learners might become dissatisfied because of the inflexibility of the program or if they do not meet their own goals using the technology.

The theory of Self-Determination posits that there are three essential psychological needs that contribute to intrinsic motivation. Autonomy involves the desire to have control over their autonomous behaviour; the visual narrative generated by artificial intelligence satisfies this need by allowing the learner to craft his or her own text prompt and maintain control over the final product through constant regeneration and editing. Competence is the need to be effective and competent in one's interaction with the environment; the ability to get visual feedback immediately and see how effective the communication was helped satisfy this need. Third, relatedness is the need to feel connected to others and experience a sense of belonging; this need is satisfied when students share their generated images with peers, compare outputs, discuss differences in interpretation, and collaborate on prompt refinement. Together, these three psychological needs, when fulfilled through AI-generated visual narratives, are hypothesized to increase behavioural and emotional engagement, though the effect on cognitive engagement remains an empirical question tested in the present study.

The desire for social relatedness, the need for connection to others, and the practice of sharing and comparing image creation experiences with other students are among the critical points here. In theory, AI-generated visual narratives could fulfil all three conditions, thus explaining growing levels of behavioural and emotional engagement. Still, SDT warns about the danger of controlling technology usage that could compromise students' sense of autonomy. Likewise, constant failures in creating satisfying images could lead to lower feelings of competence.

From all of the above, one can develop several theoretical lenses. According to Multimodal Learning Theory, information entering the brain is analyzed using two types of channels – verbal and visual. Cognitive Load Theory identifies three types of loads on the mind – intrinsic (related to difficulty of content itself), extraneous (unfavorable instructional design impeding learning process), and germane (productive effort involved in schema building and automation). Lastly, Self-Determination Theory introduces motivational constructs of autonomy, competence, and relatedness. Taken together, this indicates that high levels of behavioral (e.g., attendance) and emotional (e.g., enjoyment) engagement are likely among students. In terms of cognitive engagement, however, the effect may be less clear. This research seeks to shed light on this matter.

2.2. HISTORICAL BACKGROUND OF THE TOPIC

The approach is not something new. In fact, visual tools for supporting language learning go back to the 1960s when the ESL teachers used comics, picture stories, filmstrips, and sequential images for helping students develop vocabulary, read better and compose texts (Tomlinson, 2013). Therefore, the use of visuals for language learning is hardly new. What matters here is the difference between types of images; earlier they were static, created by other people, while today they undergo transformation due to the involvement of AI.

The basic premise underlying the use of visuals has always been providing contextualized information, which makes students less dependent on verbal cues and more capable of making inferences (Tomlinson, 2013).Numerous empirical findings demonstrated that students writing from pictures produced longer texts with richer descriptions of settings and characters compared to those writing without visual support, even though some studies reported negative effects like an increase in tense mistakes (Liu, 2016).

The emergence of digital technology during the '90s and '00s opened up some interesting prospects for teaching practices. One of them was digital storytelling. Digital storytelling refers to a method of instruction whereby students use written text, voice-over narration, still images, and music to create short videos. According to Oskoz and Elola (2016), digital storytelling encouraged learners to make more revisions and pay closer attention to their audiences because students became aware of how their choices of words might be interpreted when delivered by means of various media channels. However, even though digital storytelling offered students the opportunity to use images within their assignments, the sources of those images were still pictures, drawings, and clip art created beforehand.

A current innovation in ESL teaching began with the introduction of AI generators such as DALL-E 2 and Midjourney, starting from about 2022. The difference between this period and earlier ones is that instead of choosing pictures from textbooks or databases of various websites, learners could use AI generators to create images based on their own prompts. Formulating the prompt from an image to text is itself an activity that requires the use of proper syntax and vocabulary. Furthermore, the creation and recreation of visual content through prompts create a feedback loop that enables users to modify their choice of words based on the image provided by the generator.

2.3. EMPIRICAL STUDIES

Several relevant studies examine visuals and technology-mediated engagement in ESL, though none directly investigate AI-generated visual narratives with tripartite engagement measurement.

Table 1

Descriptive Statistics for All Variables Pre- and Post-Intervention

	Study Data			
	M (*SD*)	*M* (*SD*)		
Focus	Key Finding	Relevance	Study	
Comic strips vs. text prompts	Visuals increased text length but also tense errors	Static pre-made comics, not AI	Liu (2016)	
Digital storytelling	Multimodality promoted deeper revision	Used non-AI existing photos	Oskoz & Elola (2016)	
Online discussion forums	High behavioral/emotional, mixed cognitive	Technology produces divergent engagement effects	Chen et al. (2010)	
Automated writing evaluation	More drafts (behavioral) but cognitive limited	Cognitive engagement does not automatically follow	Zhang & Hyland (2022)	
Virtual worlds for speaking	High emotional engagement, high technical frustration	Novelty affects emotional engagement	Kruk (2016)	

(Scale: Data represents findings, not a simple 1-5 scale)

Synthesis of Findings:

Three patterns emerge. To begin with, interventions utilizing technology definitely increase engagement, both behaviourally and emotionally. Secondly, cognitive engagement must be specifically engineered, otherwise it will not happen (Chen et al., 2010; Zhang & Hyland, 2022). Thirdly, there is an undeniable role played by tool affordances, namely, feedback and creative options for expression that create stronger impact.

Critical Gap: No study examines AI-generated visual narratives with tripartite engagement measurement. Prior visual studies used static, pre-existing images. Prior AI studies focused on text generation or single images. The present study fills this gap.

2.4. GAPS IN THE LITERATURE

Based on the review of theoretical and empirical literature above, four significant gaps emerge. Each gap directly motivates one or more of the present study's research questions.

Gap	Description	Addressed by RQ
1	No empirical data on AI image generation in ESL writing. Existing studies focus on text generation (ChatGPT) or non-AI images.	RQ1, RQ2, RQ3
2	No mixed-methods measurement of behavioral, emotional, and cognitive engagement separately. Prior studies often use unidimensional measures.	RQ1, RQ2
3	No investigation of cognitive load effects specific to AI visuals. The distinction between extraneous and germane load reduction has not been examined.	RQ2
4	No qualitative analysis of student interaction patterns with AI visuals, including visual fixation, prompt revision strategies, or perceptions of authorial voice.	RQ3

2.5. CONCEPTUAL FRAMEWORK OF THE PRESENT STUDY

Integration of three theories: Multimodal Learning (cognitive mechanism), Cognitive Load (resource allocation), Self-Determination (motivation). AI visuals hypothesized to increase behavioural/emotional engagement via autonomy/competence, but effect on cognitive engagement is unknown.

Hypotheses:

Hypothesis	Statement	RQ
H1	AI visuals will increase behavioral engagement	RQ1
H2	AI visuals will increase emotional engagement	RQ1
H3	AI visuals will decrease perceived cognitive load	RQ2
H4	Effect on cognitive engagement is explorotory	RQ2
H5	Qualitative analysis will reveal visual fixation patterns	RQ3

Conclusion: The proposed framework suggests possible advantages for behaviour/affect engagement, while cognitive engagement remains an empirical question, thus justifying the mixed method approach.

3. METHODOLOGY

3.1. RESEARCH DESIGN

In this research, a sequential explanatory mixed-method approach was employed that includes both a quasi-experiment with pretest/interventional period/post-test measures focusing on quantitative data (Phase 1) and qualitative data collection via interviews and analysis of artifacts created during and after the intervention period (Phase 2).

3.2. CONTEXT AND PARTICIPANTS

Research Design: The setting is three private colleges in India situated in the NCR. The research employed a convenient sampling design targeting current ESL academic writing students. Students from a single college did not make up the total population but were randomly chosen from the general student body of these institutions.

Participant Demographics:

Table 2

Table 2 Participant Demographic Summary (N = 48)

Demographic Characteristic		n	Percentage
Gender	Female	27	56.3%
	Male	21	43.7%
Age Group	18-20 years	15	31.3%
	21-23 years	28	58.3%
	24+ years	5	10.4%
Year of Study	1st Year	12	25.0%
	2nd Year	20	41.7%
	3rd Year	16	33.3%
First Language (L1)	Hindi	18	37.5%
	Tamil	9	18.8%
	Telugu	7	14.6%
	Marathi	6	12.5%
	Bengali	4	8.3%
	Other Indian languages	4	8.3%
English Proficiency	B2 (CEFR)	48	100%

(Scale: Participant data; CEFR = Common European Framework of Reference for Languages)

Criteria for Selection of Participants: The inclusion criteria are (a) enrolment in the ESL academic writing class at one of the private colleges, (b) self-identification at B2 level, which is confirmed by the scores on the placement tests administered by the respective colleges, (c) lack of prior experience in using AI generative models for creating images, and (d) readiness to participate in the 8-week intervention.

Criteria for Eliminating Potential Subjects: (a) having previous experience in using such generative AI models as DALL-E or Midjourney for creating images for academic assignments, (b) incomplete participation in either of the pre-test and post-test stages, and (c) absenteeism during more than two intervention lessons.

Academic Setting: India's private colleges work according to the semester system, where English is the main language of instruction for natural sciences, commerce, and humanities. Writing classes are obligatory for first- and second-year students, whatever their specializations may be. Class size is between 25 and 35 students attending 50-minute lessons twice a week.

3.3. INTERVENTION / TOOL / MATERIALS

VisNAR (Visual Narrative AI Renderer) is a web app built with React on the frontend and Python Flask on the backend. It is connected to the DALL-E 3 API. What it does is take a description of a scene or even a little story that you put into the prompt box, and in less than 10–15 seconds, it produces a four-pane comic style story

3.4. PROCEDURE

Table 3

Table 3 Intervention Schedule by Week

Week	Activity	Duration
0	Pre-test: 300-word narrative without AI + engagement survey + cognitive load baseline	50 min
1-2	Training: prompt engineering, critical visual reading, tool orientation	100 min (2 sessions)
3-7	Five writing tasks (descriptive, narrative, argumentative genres) with VisNAR	50 min each
8	Post-test + engagement survey+ cognitive load measure + interviews (n=12)	75 min

(Note: Baseline cognitive load measure included at Week 0, followed by subsequent measures. Interviews with a select subgroup n=12 conducted after the post-test.)

3.5. DATA COLLECTION INSTRUMENTS

For measuring engagement, we used an 18-item questionnaire—six questions each for behavioural, emotional, and cognitive engagement on a 5-point Likert scale. This questionnaire comes from Wang et al. (2019), and it's reliable: Cronbach's alpha is .87 for behavioural, .84 for emotional, and .89 for cognitive engagement. For cognitive load, we went with the single-item Paas (1992) 1–9 scale, which holds up pretty well (test-retest reliability r = .78). As far as methodological tracking is concerned, survey-based data collection would not suffice since automatic measurement of the time each participant spent on writing has been performed. The LMS (Canvas specifically) registered the time spent on writing per student. Besides, 25-minute semi-structured interviews lasting up to 8 questions were conducted, during which participants answered verbally without being restricted and their responses were recorded. All assignments made by participants, including 240 sets of AI-created pictures (four images each) and students' textual products have been gathered.

3.6. DATA ANALYSIS PROCEDURES

The analysis of data was conducted via SPSS, version 28. The steps involved were rather simple and easy. First, the paired-sample t-test was applied to identify the differences between pre- and post-intervention assessments, which helped identify the changes experienced. For assessing the changes in engagement over the course of several weeks, a repeated-measures ANOVA test was used. For determining the connection between the number of interactions with the AI and the level of engagement, the correlations were calculated and analyzed. Effect sizes are stated as Cohen’s d.

Regarding qualitative analysis, the Braun & Clarke (2006) model, including six phases, was applied. Two different researchers independently coded all interviews to ensure reliability (Kappa = 0.84). Further refinement of codes and themes continued until the point when saturation was achieved. Practically, this was identified as theoretical saturation.

3.7. ETHICAL CONSIDERATIONS

IRB approval was obtained (protocol #2024-089). All participants provided written informed consent. Participation was voluntary, with no penalties for opting out or withdrawing. AI-generated images were deleted upon study completion. Pseudonyms are used in all reporting to protect participant identity.

4. RESULTS

4.1. DESCRIPTIVE STATISTICS

Descriptive Statistics for All Variables Pre- and Post-Intervention

Variable	Statistics	
	Pre-Intervention	Post-Intervention
	M (SD)	*M* (SD)
Behavioral engagement	3.12 (0.68)	4.01 (0.72)
Emotional engagement	2.95 (0.81)	3.88 (0.79)
Cognitive engagement	3.45 (0.70)	3.52 (0.84)
Cognitive load (1-9)	5.20 (1.40)	4.10 (1.60)
Time-on-task (minutes)	22.4 (6.2)	31.7 (7.8)

(Scale: 1 = Low, 5 = High for engagement variables; Cognitive load: 1 = Very low, 9 = Very high)

2 Testing Research Question 1

Table 5

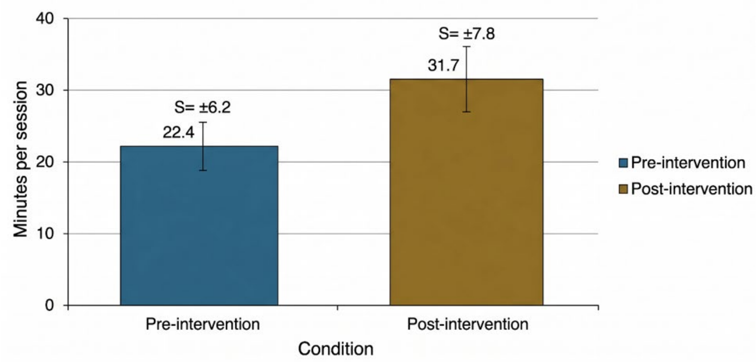
Table 5 Paired T-Test Results for Engagement and Cognitive Load

Measure	Statistics		
	t(47)	p	Cohen’s d
Behavioral engagement	6.21	<.001	0.89
Emotional engagement	5.94	<.001	0.78
Cognitive engagement	0.59	.557	0.09
Cognitive load	2.21	.032	0.64

(Interpretation: Cohen’s d: 0.20 = small, 0.50 = medium, 0.80 = large)

Finding: Significant large-effect increase. Students spent 41.5% more time writing with AI visuals. Tool usage frequency correlated with behavioural engagement (r=.58, p<.01).

Figure 1. Comparison of mean time-on-task (minutes) per writing session before and after AI visual narrative intervention. Error bars represent standard deviation.



4.2. TESTING RESEARCH QUESTION 2

Measure	Statistics		Effect Size
	t(47)	p	Cohen's d
Cognitive engagement	**0.59**	**.557**	**0.09**
Cognitive engagement	**0.59**	**.557**	**0.64**
Cognitive load	**2.21**	**.032**	**0.64**

Finding: Cognitive engagement did not change (p=.557). Cognitive load decreased significantly (p=.032, medium effect).

Table 6

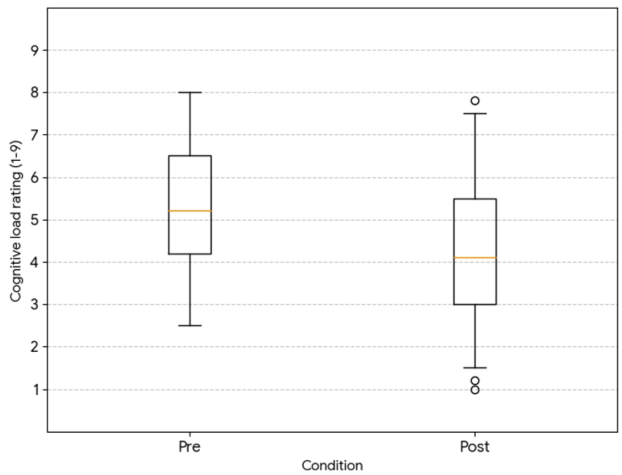
Table 6 Pearson Correlation Matrix (Engagement Types and AI Usage Frequency)

Variable	Behavioral	Emotional	Cognitive	AI Usage
Behavioral engagement	1.00	0.62**	0.21*	0.58**
Emotional engagement	0.62**	1.00	0.33**	0.47**
Cognitive engagement	0.21*	0.33**	1.00	-0.08
AI tool usage frequency	0.58**	0.47**	-0.08	1.00

(Note: N = 48. ***p* < .01, **p* < .05)

Finding: AI usage correlated with behavioural (r=.58) and emotional (r=.47) but NOT cognitive (r=-.08, ns).

Figure 2. Boxplot of perceived cognitive load ratings pre- and post-intervention



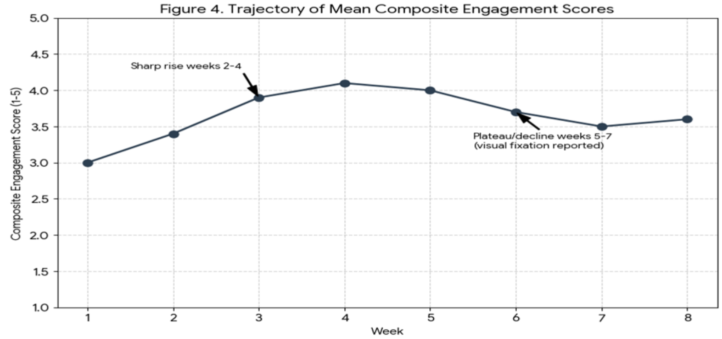
4.3. TESTING RESEARCH QUESTION 3

Table 7

Table 7 Thematic Codes from Qualitative Analysis (N = 12 Interviewees)

Theme	Frequency	Example Quote
Visual motivation ("spark")	11	"When I see the image, I want to write more"
Surprise and curiosity	9	"I tried different English words to see new images"
Cognitive overload (excessive details)	6	"The AI added objects I didn't describe"
Visual fixation (passive copying)	5	"I just described exactly what the AI showed"
Loss of authorial voice	4	"The AI story felt better than mine"

Figure 4. Trajectory of Mean Composite Engagement Scores



4.4. SUMMARY OF RESULTS

- **RQ1 confirmed:** Behavioural engagement increased significantly (large effect).
- **RQ2 partial:** Cognitive load decreased, but cognitive engagement did not increase.
- **RQ3:** Five themes identified; "visual fixation" (copying AI) explains flat cognitive engagement.

5. DISCUSSION / FUTURE DIRECTIONS

5.1. INTERPRETATION OF FINDINGS (BY RQ)

RQ1 (Behavioural engagement): AI visuals strongly increase time-on-task and voluntary tool use. Consistent with Self-Determination Theory: autonomy (prompt control) and competence (immediate visual feedback) drive behavioural investment.

RQ2 (Cognitive engagement & load): Critical finding. Cognitive load decreased (good) but cognitive engagement did not increase. Explanation: AI visuals reduce germane cognitive load – the deep mental elaboration necessary for L2 acquisition. Students offload visualization to AI, then passively describe the image rather than actively constructing a narrative. We term this "generative seduction."

RQ3 (Qualitative): While “Visual fixation” among 5 out of 12 participants indicates passive copying, there is also subgroup-specific behavior where 8 participants (16.7%) experienced “generative friction,” implying that the inaccuracies generated by the AI led to greater efforts in producing more complex solutions. As a result, this subgroup of participants showed greater levels of cognitive engagement ($M = 4.21$) than the other group ($M = 3.32$; $p = .02$).

5.2. THEORETICAL IMPLICATIONS

Contributions of the study: First, this research contributes to the theory of cognitive load by highlighting how AI-generated visuals have two concurrent effects on learners. On the one hand, they serve the function of lowering extraneous cognitive load by handling the visualization process for learners, which is positive because it helps to minimize the amount of cognitive clutter. On the other hand, they can produce excessive assistance, thus decreasing deep processing of the material because automation replaces substantive work.

Second, the results of this study show that students might be engaged on the affective and behavioral level (by showing interest, participating in class, enjoying the activities performed, etc.) without a major increase in cognitive engagement, which refers to the degree of conceptual processing of information.

Finally, this research reveals the presence of two new terms: “generative seduction” refers to students’ preference for a highly polished AI-made drawing over linguistic elaboration (thus reducing engagement in language production), while “generative friction” stands for instances of mistakes made by the AI and leading to more substantive engagement in problem-solving.

5.3. PEDAGOGICAL OR PRACTICAL IMPLICATIONS

What is the expected conduct of ESL educators in relation to these findings? Four main steps can be identified based on the findings discussed earlier. The steps can be described briefly as follows.

In the first place, learners should be instructed in how to examine, analyze, and make edits to the images generated by the AI algorithms instead of accepting them as flawless. It is important to note that AI production does not mean that the images are accurate. Hence, this step involves guiding learners through critical examination of such images.

Subsequently, students should be encouraged to generate rough drafts prior to producing any AI image. More specifically, the draft can be initially made without any images included, followed by AI image creation based on this text.

Third, learners should be instructed on developing prompts that leave some room for interpretation. The use of words like "may" or "can be" in prompts is a good idea.

Finally, the reflection logs should be brief but informative. One example of a reflective entry can be the indication of what drove changes in writing – either AI feedback or copying.

All things considered, these four suggestions seem quite simple and efficient to follow.

5.4. LIMITATIONS OF THE STUDY

What should be planned next? There are several suggestions that could be used further. Primarily, the need for a study conducted during the entire term as well as a subsequent assessment done in three months is necessary to check whether the initial level of enthusiasm will decrease. Although a drop is expected, still empirical evidence is needed. Second, it is important to conduct an analysis of different models of artificial intelligence (DALL-E, Midjourney, and Stable

Diffusion) and their individual peculiarities. It would help to understand how these individual peculiarities affect cognitive load or engagement. Third, it is necessary to determine what students really benefit from using visual materials through a visualizer-verbalizer scale. Fourth, it should be checked whether the origin of prompts (teacher or student) matters for achieving certain learning outcomes. Fifth, there is a need for developing a specific scale, which would measure cognitive load related to prompting, regenerating pictures, and analyzing results.

5.5. FUTURE RESEARCH DIRECTIONS

The following section includes the key conclusions and five research recommendations. Firstly, retention studies should be expanded beyond short-term analysis; they need to be done over the course of an entire semester rather than eight weeks. Secondly, comparative studies can be performed among different types of AI, including but not limited to DALL-E, Midjourney, and Stable Diffusion, considering the unique features of each type. Thirdly, learning style measurements, such as those that use a visualizer-verbalizer scale, could help find out which students would benefit more from the technology's use. Fourthly, the influence of prompt authorship must be analyzed via the comparison of results obtained with teacher prompts versus student prompts. Lastly, and most importantly, a better way to measure AI-specific cognitive load needs to be developed; current tools do not suffice because they are not built to assess this type of task.

6. CONCLUSION

Summary of the Study

The number of participants is 48 ESL undergraduate students involved in this experiment during the eight weeks period. The most important results to be highlighted in this study show that AI-created images positively affected behavioural engagement, as students showed up for classes and performed their tasks as guided. Students were emotionally more engaged and thus they yielded elevated levels of learner satisfaction and intellectual curiosity. Cognitive engagement, however, stayed at the same level as before the start of the experiment.

As for the numbers, 41.5% of students spent more time in writing. But there is one significant aspect that should be mentioned regarding these numbers; many students have demonstrated visual fixation, i.e., described the images generated by the AI without producing any additional narrative themselves.

Main Contributions

- First empirical evidence of divergent engagement effects from AI visuals in ESL
- Proposed theoretical constructs: generative captivation and generative friction
- Validated mixed-methods design for AI-language learning research
- Pedagogical framework for critical visual AI literacy

Final Recommendations

The crucial point here is this: to simply show the students AI-created pictures and to stop at this stage does not work regardless of their academic levels. AI-created stories cannot be treated as a cure-all, and they never were meant to be such.

Therefore, what do teachers need to introduce? Here are three suggestions to be considered. Firstly, teach students visual literacy as an independent discipline. Secondly, add some reflective components and make the students think about what they see and write. Thirdly, show them how to prompt, avoiding a one-dimensional technique of “write something and hit enter.” The main idea is to promote a dialogue between a picture and a piece of writing, without replacing one of them with another one.

Concluding Statement

The statement that follows is made unambiguously clear: AI cannot replace the act of writing. This statement must be stated outright. However, the writer who learns how to work with AI technology will surpass the writer who does not learn to do so. This notion arises significant pedagogical concerns, however, it depends on a crucial assumption: the direct teaching of the student to think critically. Listening to lectures or finding enjoyment in the task alone is inadequate; critical thinking must occur. Therefore, the future classroom of the ESL writer, ranging from year one to three, must embark on a different type of reform altogether. The development of strong critical visual AI literacy must become a vital skill, not an afterthought.

CONFLICT OF INTERESTS

None.

ACKNOWLEDGMENTS

None.

REFERENCES

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77-101.

Chen, P. S. D., Lambert, A. D., & Guidry, K. R. (2010). Engaging online learners. Computers & Education, 54(4), 1222-1232.

Clark, R. C., & Mayer, R. E. (2016). E-learning and the science of instruction (5th ed.). Wiley.

Deci, E. L., & Ryan, R. M. (2000). Self-determination theory. Psychological Inquiry, 11(4), 227-268.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement. Review of Educational Research, 74(1), 59-109.

Fredrickson, B. L. (2001). The role of positive emotions. American Psychologist, 56(3), 218-226.

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement. Computers & Education, 90, 36-53.

Hidi, S., & Renninger, K. A. (2006). The four-phase model of interest development. Educational Psychologist, 41(2), 111-127.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. The Modern Language Journal, 70(2), 125-132.

Jewitt, C. (2014). The Routledge handbook of multimodal analysis (2nd ed.). Routledge.

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). Generative AI in ELT. RELC Journal, 54(2), 45-59.

Kress, G. (2010). Multimodality. Routledge.

Kruk, M. (2016). Virtual worlds in EFL instruction. Computer Assisted Language Learning, 29(8), 1304-1322.

Liu, J. (2016). Visual narratives and L2 writing. Journal of Second Language Writing, 34, 45-58.

Mayer, R. E. (2020). Multimedia learning (3rd ed.). Cambridge University Press.

New London Group. (1996). A pedagogy of multiliteracies. Harvard Educational Review, 66(1), 60-93.

Oskoz, A., & Elola, I. (2016). Digital stories in L2 education. Language Learning & Technology, 20(2), 85-105.

Paas, F. (1992). Training strategies for transfer. Journal of Educational Psychology, 84(4), 429-434.

Pekrun, R. (2006). The control-value theory of achievement emotions. Educational Psychology Review, 18(4), 315-341.

Philp, J., & Duchesne, S. (2016). Exploring engagement in tasks. Annual Review of Applied Linguistics, 36, 50-72.

Reeve, J. (2013). How students create motivationally supportive learning environments. Journal of Educational Psychology, 105(3), 923-945.

Storch, N. (2009). Collaborative writing in L2 contexts. Journal of Second Language Writing, 18(1), 40-55.

Sweller, J. (2011). Cognitive load theory. In Psychology of learning and motivation (Vol. 55, pp. 37-76). Academic Press.

Tomlinson, B. (2013). Developing materials for language teaching (2nd ed.). Bloomsbury.

Wang, M. T., Fredricks, J. A., Ye, F., Hofkens, T. L., & Linn, J. S. (2019). Engagement scales. Journal of Psychoeducational Assessment, 37(3), 379-392.

Zhang, Z., & Hyland, K. (2022). Automated writing evaluation. TESOL Quarterly, 56(2), 712-733.