Impact of SORA AI-Generated Storyboards on CLIL Writing Anxiety, Motivation and Enjoyment

Esen Genç

[1] Ondokuz Mayıs University, Türkiye

*Corresponding author's email: esen.genc@omu.edu.tr

*ORCID: http://doi.org/0000-0002-7700-4806

Received: 02/06/2025 Revision: 04/08/2025 Accepted: 22/08/2025 Online: 31/12/2025

ABSTRACT

The purpose of this study is to explore if OpenAI's SORA, an AI program that creates video storyboards, can better provide writing related affective outcomes than traditional hand drawn sequences in a Content and Language Integrated Learning (CLIL) situation. A total of seventy-two undergraduate participants who enrolled in an English preparation program were randomly assigned to either a control or experimental group. The experimental group utilized SORA to transfigure written prompts into a brief video, while the control group produced nine hand images in a session for thirty minutes using contextual prompts. While both groups did not result in significant pre–post changes on the main scales, the control group had a statistically significant decline in motivation, while the experimental group held steady at baseline motivation levels. The engagement scores in the SORA condition were high even though students expressed concern with time and complications designing prompts. Taken together, these results provide evidence that a generative AI to complete storyboards can impact learners' motivation and engagement during CLIL writing tasks in a positive manner because it is a positive multimodal support that uncouples cognitive capacity from the drawing technicality and provides an opportunity for learners to communicate and elaborate linguistic ideas.

Keywords: Artificial Intelligence, SORA, CLIL, EFL

Introduction

Artificial Intelligence (AI) has been known to have a big potential in processing huge amounts of data, analyzing complex relationships, and providing individualized insights in various fields. Thus, it constitutes an important and revolutionary tool for English as a foreign language (EFL) education (Ouyang & Jiao, 2021; Chen et al., 2020). The incorporation of AI into language education requires a significant transformation towards approaches that emphasize more creative, linguistic, and affective dimensions of learning (Wang et al., 2023; Roll & Wylie, 2016). Thus, being one of these skills that go hand in hand with creativity, writing is suitable for engaging in AI-driven multimedia platforms that promote learner engagement. One of these platforms, SORA, was developed by OpenAI and introduced in early 2024, is a generative AI model that was designed to transform language prompts into short video clips.

Although it was initially released mainly for artistic purposes, its potential for EFL has rapidly gained attention in terms of providing visual scaffolding that can reduce cognitive load and enhance conceptual encoding. An AI-mediated, multimodal scaffold such as SORA could be assumed to shape affective outcomes in EFL writing classes. It can empower students to decide what to visualize by producing professional-looking clips regardless of artistic skill and support peer sharing of videos. This is in line with one of Self-Determination Theory's key requirements that tasks nurturing autonomy, competence, and relatedness sustain intrinsic motivation (Deci & Ryan, 2000). The instant visual feedback from SORA may also lower anxiety by removing worries over drawing ability and time management, facilitating EFL input uptake. As Paivio(1991) argues, information encoded verbally and visually enhances learning and recall; generative video clips automatically supply the visual channel, allowing learners to concentrate cognitive resources on L2 expression. Accordingly, it can support Cognitive Load Theory (Sweller, 2010), which emphasizes the importance of differentiating intrinsic, extraneous, and germane load by cutting extraneous load and free capacity for germane load devoted to organizing content and language. These perspectives form an integrated conceptual model linking AI use, multimodal scaffolding, cognitive load, and affective learning.

In light of these assumptions, this study aims to investigate the influence of SORA on writing-related motivation, anxiety, engagement, and enjoyment while also being mindful of these assumptions. It examines SORA-based video storyboard creation in comparison to hand-drawn approaches, and the impact that the modality of visual representation has on the affective and motivational qualities of writing practices in EFL classes. A CLIL context is chosen because of the anticipation of interactions between language production and content learning, which is often at the heart of the CLIL approach. It is anticipated that meaningful interactions can be supported when learners are afforded multimodal opportunities that support cognitive and linguistic growth (Coyle et al., 2010; Villabona & Cenoz, 2022). Storytelling in EFL contexts has been previously identified as improving engagement and writing fluency because it assists learners with bridging their inner and external language representations (Ivanytska et al., 2024). At this stage, science fiction storyboards can encourage creative expression by handling the technical aspects of visual production. Students would now be able to pay more attention to ideation, coherence, and elaborating upon their language, rather than being limited by their ability to draw (Behrens et al., 2024). The affective domain of EFL, specifically in writing, is still a less examined area of CLIL research. Current research has noted that CLIL instruction using multimodal supports alleviates foreign language anxiety and promotes a greater sense of affective climate in learning. Similarly, AI-supported language instruction has been shown to promote increased writing ability and motivation among EFL students (Song & Song, 2023; López-Medina & Casado, 2024). Nonetheless, despite patterns of cognitive and performance based outcomes stemming from AI integration, a few studies have explored how different visual creation modalities influence the affective domain in writing. Additionally, there still remained a need to explore how these tools interplayed with the dual cognitive load placed on learners in CLIL environments, having to negotiate both content and language accuracy (Villabona & Cenoz, 2022).

The present study seeks to address this gap through a comparison of two distinct modes of visual storyboarding in CLIL-based foreign language writing: AI-generated video clips using

SORA and conventional hand-drawn sequences. The study serves to appraise the comparative impact of these two modalities on students' writing anxiety, motivation, enjoyment, and engagement. Through its examination of these affective measures, the study offers a more comprehensive understanding of how innovative and emerging AI tools can be used ethically and effectively to further language education. At the same time, it also provides actionable suggestions for educators who would like to respond to innovation, while balancing its pedagogical rigor taken up in EFL classrooms.

Literature review

Being defined as an analytical life set which aims to imitate real-life (Gordon, 2011), AI was first used by McCarthy et al. in their 1955 summer school research project based on the assumption that any kind of learning or other attributes of intelligence could, in principle, be done by a machine to imitate it (McCarthy et al., 1955). Today, with endless investments and technological developments, AI applications are increasing day by day in terms of performance and awareness. It has a wide range of usage fields from Siri to digital journalism, from predicting stock movements to predicting crime, from facial recognition to medical diagnostics (Holmes et al., 2019). Unsurprisingly, AI based technologies has started to be integrated in education together with other aspects of daily life. Along with the advancement of machine learning algorithms, AI applications began offering more flexible, personalized instruction opportunities by combining machine inference with human pedagogy (Lee & Lee, 2024).Today, researchers are now even discussing about AI's potential to replace "real" teachers, although most studies have revealed some concerns regarding this possibility (Chan & Tsi, 2023; Tao et al., 2019). AI's ability of assisting learners at any time (Rojas-Muñoz., et al, 2021) and improving existing teaching and learning practices by making necessary arrangements make it a valuable tool (Wang & Cheng 2021). Despite some concerns like privacy, technical issues and over reliance, the integration of AI into EFL classes offers various advantages like individualized learning and assessment that can assist students and teachers (Verdenhofa et al,. 2024).

The ability of AI to analyze environments and perform human-like tasks makes it a valuable instrument for teaching productive language skills in EFL classrooms. The newly developed generative capabilities of AI have enabled platforms such as which enables students to watch narrative content instead of creating their own storyboards. The modality supports cognitive load theory by providing visual scaffolding which helps conceptual encoding especially in CLIL contexts where students handle subject content and language requirements (Villabona & Cenoz, 2022). Therefore, the integration of AI into EFL, particularly through multimodal and generative tools, holds significant promise for enhancing learner engagement, reducing cognitive load, and fostering more effective development of language skills within content-rich instructional settings.

Writing is a complex activity that requires the coordination of several different elements, such as psychological, motivational, metacognitive, and cognitive processes (Hayes, 2012). Thus, learner psychology and emotions also have important roles in the complex processes of EFL writing, and that cognition alone is not enough to control those (Wang et al., 2024). Such psychological processes and emotional experiences, especially if the cognitive demands of the tasks impact them, may contribute to divergent task performance and learning accomplishment

(Li et al., 2021). As a significant emotional state, foreign language anxiety, which can be defined as the feeling of tension generally related to foreign language learning environments (MacIntyre & Gardner, 1991), may affect the learners' creativity. Motivation is another important factor in EFL classes, and it stands out as a fascinating concept for teachers and researchers because it can be developed through proper social circumstances (Noels, 2001). Research shows that motivation to learn a second language equals or surpasses language aptitude as a critical factor for successful language acquisition (Gardner, 1985).

As there is a growing interest towards AI-driven platforms in EFL classes, the discussion about their effects on emotional aspects is of great importance. Up to now, there have been some studies that focus on the subject. The affective benefits of AI are highlighted by Mohammed and Khalid (2025), who found that AI-generated feedback improved learners' motivation, emotional well-being, and Foreign Language Peace of Mind (FLPoM),a state of reduced anxiety and increased emotional stability grounded in Krashen's (1982) affective filter hypothesis. This explains the dual cognitive-emotional scaffolding provided by intelligent systems in regard to real-time responses to learner input. Students who availed of AI-based story generators had better creativity and linguistic elaboration according to Woo et al. (2024) in their narrative writing tasks. Song and Song (2023) found that EFL students with AI-assisted feedback tools had better writing skills and greater motivation. While numerous studies document cognitive gains from AI feedback (e.g., Song & Song, 2023) or reduced anxiety through multimodal CLIL materials (Villabona & Cenoz, 2022), three gaps remain. First, a modality-contrast gap: prior research rarely isolates the effect of the method of visual creation (AI video vs. student drawing). Second, a motivational-trajectory gap: findings regarding motivation are mixed, some report short-term boosts (Behrens et al., 2024) whereas others show novelty effects fading quickly. Third, an integrated-affect gap: no study, to my knowledge, simultaneously examines anxiety, enjoyment, and motivation stability within one design that controls for time-on-task. Conflicting evidence underscores the need for a study that manipulates storyboard modality, tracks multiple affective dimensions over time, and is situated in a CLIL context where cognitive load is already high. Yet none of these studies contrasted AI-generated videos with learner-drawn storyboards, leaving open the question of how creation modality interacts with the affective filter, motivational stability, and enjoyment.

In sum, the trajectory of AI in education, from static rule-based tutors to generative, multimodal systems, reveals an expanding capacity to support not only cognitive but also affective and creative aspects of language learning. Especially in CLIL environments, tools like SORA offer promising means to lower cognitive load, enhance learner motivation, and foster authentic engagement with language through visually enriched storytelling.

At this point, the present study aims to contribute these three points. First, it clarifies the modality effect on motivation stability by showing that SORA sustains baseline motivation while hand drawing does not, thereby resolving inconsistent past results. Second, it presents an integrated affective profile by measuring anxiety, enjoyment, and engagement concurrently in CLIL writing. Third, it offers an applied scaffold-design blueprint, demonstrating how prompt training and pacing operationalize Dual Coding and Cognitive Load principles. By directly comparing AI-generated video versus hand-drawn storyboards, this work addresses a variable in multimodal language learning and extends CLIL-AI research into the affective zone where

motivation and anxiety jointly predict sustained EFL writing effort.

Research Questions

In light of these assumptions, this study attempts to compare SORA-based video storyboard creation with traditional hand-drawing practices. Based on the interest in this situation, the following research questions are formed:

1. Does using SORA to generate video storyboards reduce writing anxiety more effectively than drawing images?
2. Are there differences in learner motivation and enjoyment between groups?
3. How does user engagement during the intervention relate to post-intervention outcomes?

Methods

Pedagogical Setting & Participants

The target population consisted of undergraduate students registered in an English preparatory program at a state university. Seventy-two volunteer students were recruited using a convenience sample (Fraenkel & Wallen, 2006) and randomly assigned to treatment conditions (i.e., experimental, control) by way of a random number generator. Each participant was assigned a random ID, and after assignment, 36 participants were randomly assigned to the experimental and 36 participants were assigned to the control conditions through random shuffling. Participants in each condition formed subgroups of five, resulting in a total of 12 small groups per condition. The inclusion criteria for participants were: (a) current enrollment as a university student, (b) age 18 years or older, (c) having basic computer skills and (d) written informed consent was provided. Participants were excluded from the study if informed consent was not provided or if the participant failed to attend any of their assigned intervention sessions. Randomization occurred at the individual level using random number generation in Excel.

Design of the Study

This study utilized a quasi-experimental pretest-posttest design with a control group to analyze the effects of AI-generated versus hand-drawn storyboards on affective writing outcomes in a CLIL classroom. Although participants were assigned randomly to a collaborative storyboard group versus a control group, the overall sampling was non-random and was based on convenience and fits with a quasi-experimental design (Creswell, 2014). A quasi-experimental design enables strong comparisons between a treatment and control group when true experimental manipulation of all extraneous variables cannot be realistically implemented. To keep conditions constant in all five sessions, both groups used standardized written lesson plans, and the session was co-facilitated with two researchers who engaged in a joint-designed training session prior to implementation. A fidelity checklist was used after each session to document adherence to instructional protocols (e.g., time allotted, instructions for the task, and use of either the SORA or the drawing materials). Peer debriefing of both facilitators occurred weekly to address any convergences of procedures.

Data collection & analysis

Five established self-report scales were administered:

1. **Writing Anxiety Scale**

   To evaluate the students' anxiety related to writing, Cheng's (2004) *Second Language Writing Anxiety Inventory* (SLWAI) was used. The SLWAI is comprised of 27 items that tap into the cognitive, somatic, and behavioral indicators of anxiety experienced while writing. Each item is evaluated on a 5-point Likert scale (1 = strongly disagree; 5 = strongly agree), with higher scores representing higher anxiety. The original study reported excellent internal reliability ($\alpha = .91$). In the current study, the SLWAI again demonstrated strong reliability (Cronbach's $\alpha = .89$).

2. **Learner Motivation Scale**
   A 12-item scale developed by Ardasheva, Tong, and Tretter (2012) was used to measure learners' intrinsic and extrinsic motivation in a language learning context. The original measure had acceptable overall reliability ($\alpha = .80$), although the alpha coefficients for some subscales were between .58 and .74. In the current study, the instrument demonstrated acceptable reliability, with a Cronbach's alpha of .83 suggesting good internal consistency for the items in the current sample.

3. **Enjoyment Scale**
   To measure enjoyment when completing tasks, we employed the 21-item *Foreign Language Enjoyment Scale* developed by Dewaele and MacIntyre (2014). The items refer to different facets of positive affect regarding language learning, notably fun, interest, and social enjoyment with other learners of the language. The original research exploring the scale reported an alpha of .86. The dataset in this study revealed a similar level of internal consistency, with a Cronbach's alpha of .87.

4. **User Engagement Scale**
   The 12-item short form of the *User Engagement Scale* (O'Brien, Cairns, & Hall, 2018) was used to evaluate learners' perceived attention, usability, and reward while engaging with the writing tasks. Previous research reported subscale reliabilities ranging from .70 to .90. The present study found a comparable level of internal consistency, with a Cronbach's alpha of .85.

5. **Demographic Questionnaire**
   A short demographic form was used to collect participants' age, gender, and previous experience with artificial intelligence tools or hand-drawn sketching.

All scales have demonstrated good reliability and validity in previous EFL contexts (Cheng, 2004; Ardasheva et al., 2012; Dewaele & MacIntyre, 2014; O'Brien et al., 2018). In this study, internal consistency was checked via Cronbach's alpha; only scales with $\alpha \geq .75$ were retained.

Procedure

The intervention lasted for five 60-minute sessions over two weeks, which may not have been sufficient to generate long-term or deeply internalized affective changes. While short-term engagement and immediate responses were captured, future research should extend the intervention period to assess longer-term shifts in writing-related anxiety and motivation.
Each of the five sessions focused on a different writing prompt scenario, totaling nine prompts (two visualized per session; one for warm-up and one for the main task). These scenarios were designed collaboratively by two applied linguistics researchers and one EFL instructor, and were drawn from common themes in university-level writing (e.g., personal experience, process description, problem-solution). Prompts were:

- Matched for genre, linguistic complexity, and cognitive demand,
- Reviewed by a panel of two additional EFL experts for content equivalence across sessions and groups, and
- Piloted with a separate class (n=18) one semester earlier to confirm appropriateness and clarity.

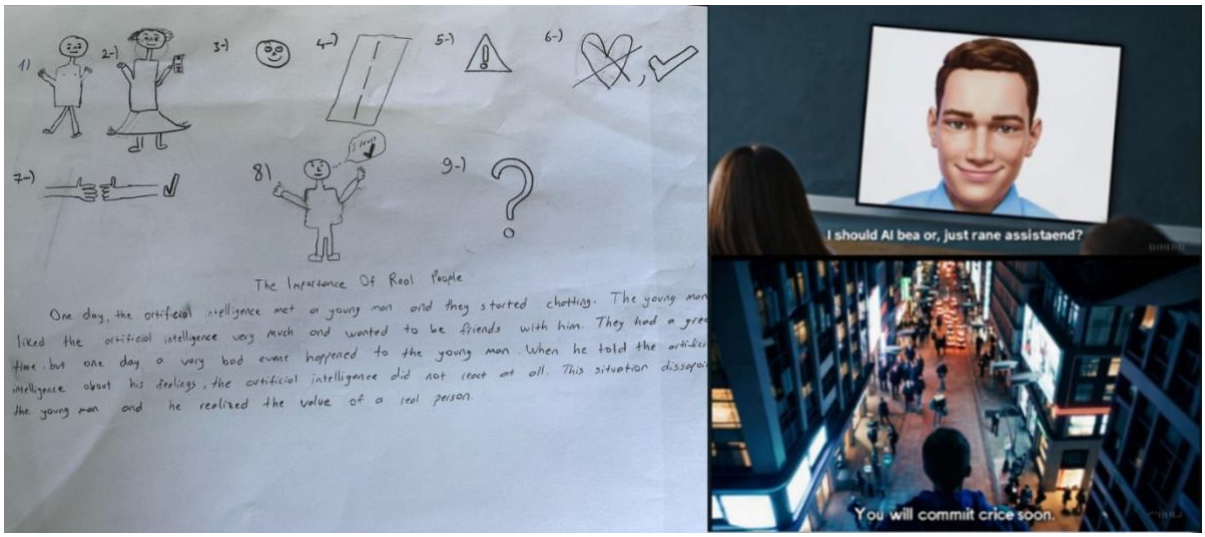
Pre-Intervention (Week 1)

In the first week, all participants gathered face-to-face in their regular classrooms for an introductory session. After a concise overview of the study's aims, each person signed an informed consent form (Ethics approval OMU: 2025-196) and then completed a demographic questionnaire along with three baseline measures: Writing Anxiety, Motivation, and Enjoyment.

Intervention (Weeks 2–3)

Over the next two weeks, both the experimental and control groups met for five 60-minute sessions designed to foster storyboard creation. Every team in both groups were given a sheet in which nine scenarios based in the theme of AI were written. The students were provided with a degree of autonomy to choose a scenario they preferred (See Supplementary File 1**).**In the experimental condition, sub teams used SORA to generate nine short videos from written prompts and then wove these clips into a coherent video storyboard. In contrast, the control group worked with pencil and paper, crafting nine hand-sketched images to assemble a traditional storyboard. In both groups, after creating their storyboards for 30 minutes, they were given another 30 minutes more to write a 180-220 words story based on their story boards (Figure 1). Notably, at the midpoint of every session, we administered the User Engagement Scale to capture in-task engagement, while research assistants quietly observed each group ensuring procedural fidelity and answering any questions that arose.

Figure 1.
Student created storyboard versus SORA generated storyboard



Post-Intervention (Week 4)

During the fourth week, the participants came back to the classroom for post-test assessments. For participants that missed the classroom assessments, I sent the same surveys using Google forms. The participants then completed the post interventions versions of the Writing Anxiety, Motivation, and Enjoyment scales to directly compare the results back to the pre-intervention baseline.

Data Analysis

Data from semi-structured interviews were analyzed via thematic content analysis. Qualitative analysis followed a six-phase method of thematic analysis (Braun & Clarke, 2006). Two coders independently reviewed the transcripts and identified initial codes inductively and then met to group them into themes. Inter-coder reliability was analyzed using Cohen's Kappa ($\kappa = .84$), demonstrating substantial agreement. To begin, an inductively developed coding schema was embodied from five randomly chosen transcripts. Two independent coders employed the schema on the entire dataset. As before, inter-rater reliability was analyzed using Cohen's Kappa, and yielded $\kappa = .84$. A discussion process resolved any coding discrepancies as the coders reconciled the code-frequency tables. The thematic analysis process was completed manually. While qualitative analytic software (e.g. NVivo) was used in consideration, both the dataset's size and scope warranted a sufficiently reliable manual coding process. Future research may benefit for-going digital in-process auditability and transparency.

The total sample consisted of 72 participants (36 per group). While this size is adequate for detecting medium to large effects, no a priori power analysis was conducted. Post hoc calculations based on the observed effect size for motivation (Cohen's $d = .465$) indicate that the study was powered to detect medium effects, but underpowered for small effects. This limitation is acknowledged in the discussion section. Quantitative data were analyzed using SPSS 26, to increase analytical rigor, future replications should consider using ANCOVA or mixed-design ANOVA to control for baseline differences and more precisely assess interaction

effects over time. Internal consistency for each scale was evaluated with Cronbach's alpha, with $\alpha \geq .75$ considered acceptable. And then, Shapiro–Wilk tests were applied to verify normality assumptions (Field, 2013). To examine within-group changes, paired-samples t-tests compared pre- and post-test scores for each condition, while between-group effects were assessed by calculating gain scores (post – pre) and subjecting those to independent samples t-tests. Relationships between mid-session engagement and outcome changes were explored using Pearson's r. Finally, effect sizes were calculated as Cohen's d (small = .20, medium = .50, large = .80; Cohen, 1988), and statistical significance was set at $p < .05$.-high

Ethical Considerations

The study protocol was reviewed and approved by Ondokuz Mayıs University Social & Humanities Ethics Committee (Approval No. OMU: 2025-196). Participation was voluntary, with the right to withdraw at any time. Data were anonymized, stored on password-protected drives for two years, and then securely destroyed. No personal identifiers were linked to scale responses, ensuring confidentiality throughout.

Results/Findings and discussion

Cronbach's alpha was used to check the internal consistency of the scales used in the study. The engagement scale demonstrated strong internal consistency with a reliability coefficient of .821. Writing anxiety was also considered reliable with a coefficient of .803; the motivation scale was acceptable with a value of .753; enjoyment had the highest reliability with a coefficient of .876. Overall, the results indicate that all 4 scales were reliable measurement tools for measuring the identified psychological variables.

Prior to completing additional statistical tests, the Shapiro-Wilk test was used to check the assumption of a normal distribution of data. The results of the analysis indicated that anxiety, motivation, and enjoyment data were normally distributed both before and after the intervention (p values were greater than .05). Based on all variables showing non-normality, and an adequate sample size, it was decided to proceed with parametric testing.

Results within Groups

Table 1 presents the paired samples t-test results that compared pre-test and post-test scores for each group. In the control group, there was a statistically significant decrease in motivation. The mean difference was −.345 with a standard deviation of .919, and the result was significant at the $p = .031$ level. The effect size was moderate (Cohen's $d = -.375$). In contrast, anxiety and enjoyment did not show any meaningful changes in this group.

Table1.

Paired Samples t-Test Results for Anxiety, Motivation, and Enjoyment in Control and Experimental Groups

Group	Variable	M Difference	SD	*t*	*p*	Cohen's *d*
Control	Anxiety (Post–Pre)	.021	.768	.168	.868	.028
	Motivation (Post–Pre)	−.345	.919	−2.252	.031*	−.375

	Enjoyment (Post–Pre)	−.004	1.013	−.024	.981	−.004
Experimental	Anxiety (Post–Pre)	.035	.569	.373	.712	.062
	Motivation (Post–Pre)	.000	.505	.000	1.000	.000
	Enjoyment (Post–Pre)	.066	.619	.641	.526	.107

*p < .05

In the experimental group, none of the variables showed statistically significant changes between the pre-test and the post-test. Anxiety increased slightly, but the change was not significant (p = .712). Motivation stayed exactly the same, and enjoyment increased slightly, though not significantly either. These results may suggest that the intervention helped maintain student motivation and slightly improved enjoyment, but the effects were not strong enough to reach statistical significance.

Results between Groups

Independent samples t-tests were conducted to compare the amount of change between the control and experimental groups. This was done by calculating gain scores, which reflect the difference between post-test and pre-test results. Table 2 displays these comparisons.

Table 2.

Independent Samples t-Test Comparing Control and Experimental Groups on Gain Scores

Variable	Group	M Gain	SD	*F*	*t*	*p*	Cohen's *d*
Anxiety	Control (n = 36)	.021	.768	4.12*	−.087	.931	−.021
	Experimental (n = 36)	.035	.569				
Motivation	Control (n = 36)	−.345	.919	19.13**	−1.974	.052	−.465
	Experimental (n = 36)	.000	.505				
Enjoyment	Control (n = 36)	−.004	1.013	1.50**	−.354	.724	−.084
	Experimental (n = 36)	.066	.619				

*p < .05. **p < .01

As seen in Table 2, there were no statistically significant differences between the two groups in terms of gain scores for anxiety, motivation, or enjoyment. For anxiety, the control group and experimental group had very similar average gains, and the difference was negligible (p = .931). For motivation, the control group showed a decrease, while the experimental group showed no change. While this difference was approaching significance (p = .052), it did not exceed the standard threshold. Nevertheless, it had a moderate effect size (Cohen's d = −.465), which could indicate an important trend. There was a small difference in gain scores for enjoyment, but it was not statistically significant (p = .724), and the effect size was negligible. Overall, these data suggest that there were no statistically significant changes in outcome measures from the intervention. However, the trend in the motivation data may warrant additional exploration in future studies, especially considering the control group displayed a decline while the experimental group maintained motivation.

Next, I examined the engagement score of experimental group during the intervention. The experimental group of 36 students reported relatively high engagement during the Sora-based English lesson (Figure 2)

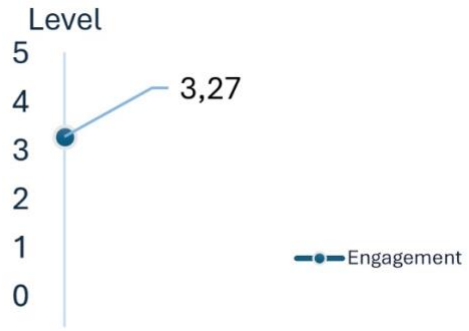


Figure 2. Engagement Level During SORA Lesson

As displayed in Figure 1, the engagement scores ranged from a minimum of 1.08 to a maximum of 4.08 on the scale used, indicating some variability in student responses. The mean engagement score was approximately 3.27 (SD = .56), suggesting that, on average, students were actively involved and interested in the lesson (Figure 1).

Content Analysis of Interview Responses

To clarify the reasons behind the quantitative results which demonstrated no significant differences between the groups in writing anxiety, motivation, or enjoyment, although there was a trend in which the experimental group (SORA) appeared to sustain motivation while the control group (drawing) appeared to have decreased motivation, semi-structured interviews were administered to 20 participants in each group. While the study had included approximately 36 participants in each group, only 20 in each group agreed to respond to the interview. An in-depth qualitative content analysis was conducted by marking responses for affective factors (anxiety, motivation, enjoyment, engagement) along with practical factors (e.g. challenges, preferences), and by synthesizing responses to contextualize the quantitative findings. The analysis did allow the emergence of themes that were distinguishable across the groups, to better understand the learner experiences which shaped the findings. The interview approach included semi-structured questions, and length varied for each participant; some provided detailed answers, nevertheless, the analysis was limited to what participants explicitly recognized when discussing the themes. One participant in the experimental condition was not available and did not answer questions which were relevant to the study. Therefore the body of inquiry was finished with a sample of 19 for the experimental group in the analysis phase.

Experimental Group (SORA): Key Themes

The high frequency of enjoyment (17/19) and motivation (16/19) reflects the novelty of SORA, aligning with the quantitative trend of maintained motivation (M gain = .000, p = 1.000). Challenges were noted by over half the participants (10/19), particularly time and prompt issues, which may explain the lack of significant changes in anxiety (p = .712) or enjoyment (p

= .526). Anxiety was rarely mentioned, consistent with the non-significant anxiety change (M gain = .035, p = .712). The absent participant's response was excluded, reducing the denominator to 19.

Table 3.

Experimental Group's Code Frequency Table

Code	Description	Frequency (n=19)
Enjoyment	Expressed enjoyment, fun, or positive feelings about the SORA writing task.	17
Motivation	Indicated feeling motivated or eager to complete the task, often due to curiosity about AI outputs.	16
Engagement	Reported feeling engaged, involved, or curious during the storyboard creation.	15
Visualization Support	Noted that SORA helped visualize the story, aiding the writing process.	14
Challenges (Time/Prompts)	Mentioned challenges like time constraints or difficulty crafting precise prompts.	10
Lack of Motivation/Engagement	Expressed disinterest, lack of motivation, or limited engagement (e.g., due to group work or AI skepticism).	4
Anxiety	Mentioned anxiety or stress related to the task (e.g., prompt clarity).	2

SORA group participants, in particular, discussed their experience as enjoyable, creative, and motivating due in part to the fact that it was novel to use AI to visualize their stories. Many were compelled by the prospect of seeing their written prompts in video form and made comments like, "It was a new experience for me, I found it creative and interesting," and "Seeing the results of prompts made up of a couple of sentences made me really excited to tinker around with SORA more." This excitement, in parallel to the quantitative finding that motivation remained consistent (M gain = .000, SD = .505, p = 1.000), indicated that the curiosity and interactivity offered by SORA maintained learners' motivation throughout the process, which observation data suggests was declining for the control group. The anticipation in waiting for the results of prompts generated by AI enhanced the writing process and fostered a more dynamic and rewarding experience.

Another key theme emerging from the data was the concept of engagement. A number of participants commented about active engagement potentially driven by SORA's immediate visual feedback. For example, one participant stated, "I felt very engaged because it was holding my attention and it made me feel more engaged and part of the process." The participants' mean engagement score was 3.27 (SD = .56) for the full experimental group (n = 36), so evidence suggested generally high engagement in this group. However, variability in scores (1.08 to 4.08) and other comments like "Since we were given such a short amount of time to come up with

prompts and we were in front of a group in a class environment, I do not feel like I was fully engaged with SORA" suggest that engagement could also be thought of as moderated by external factors, such as the availability of time and the group context, impacted the overall engagement in the intervention.

Time and prompt clarity challenges emerged as critical obstacles. Participants expressed that their limited time inhibited their ability to fully utilize SORA's capabilities, with one participant stating, "Time definitely did affect my experience, I would have loved to have a large amount of time in my hands." Additionally, developing precise prompts to get the visuals they wanted was a struggle for some, saying, "Sometimes it was challenging to find prompts that exactly matched my ideas." These practical difficulties may have influenced the finding of no significant changes in anxiety (M gain = .035, SD = .569, p = .712) and enjoyment (M gain = .066, SD = .619, p = .526), as the captivating experience of SORA was somewhat tempered due to imposed logistical constraints on the experience.

SORA's potential to facilitate visualization was well-regarded, especially by the majority of participants who said it enhanced their writing process by "bringing intangible ideas into tangible form." Participants' perceptions that "yes, it helped me visualize my story... using SORA to visualize stories is better" reflect this value. Some, however, mentioned differences between their mental images and the outputs from SORA, as in, "it helped but sometimes it was just close to my imagine." Although this visualization support very likely helped sustain motivation and small increases in enjoyment, making it possible to complete the major writing task, illustrative outputs that differed from participants' expectations may have limited on the overall influence of the intervention, consistent with little change in significance in the quantitative results.

Control Group (Drawing): Key Themes

Enjoyment (15/20) was common, which reflects the hands-on appeal of drawing, but occurred less often when compared to SORA, which was in line with no significant change in enjoyment (M gain = -0.004, p = 0.981). Motivation (12/20) and engagement (11/20) scored lower than SORA, and the significant decrease (M gain = -0.345, p = 0.031) in motivation is consistent with those indicators. Challenges, such as time or drawing skills (9/20) likely contributed to the decrease in motivation since some participants expressed frustration with the drawing process, but anxiety scores were nearly absent, suggesting a non-significant change in anxiety in this group as well (M gain = 0.021, p = 0.868).

Table 4.
Control Group's Code Frequency Table

Code	Description	Frequency (n=20)
Enjoyment	Expressed enjoyment, fun, or positive feelings about the drawing task.	15
Motivation	Indicated feeling motivated to complete the task,	12

Code	Description	Frequency (n=20)
	often due to creative expression or group work.	
Engagement	Reported feeling engaged or focused during the storyboard creation.	11
Organization Support	Noted that drawing helped organize thoughts or structure the story.	10
Challenges (Time/Skills)	Mentioned challenges like time constraints or lack of drawing skills.	9
Lack of Motivation/Engagement	Expressed disinterest, lack of motivation, or boredom (e.g., due to drawing difficulty or disliking writing).	6
Anxiety	Mentioned anxiety or stress related to the task (e.g., time pressure).	1

In the control group, participants expressed appreciation for the drawing activity for being fun and building their confidence, citing the value of meaningful and practical creativity. For instance, participants stated, "It was very enjoyable because I drew it according to my imagination and it was fun," and "It made me feel more confident and motivated." The enjoyment expressed was consistent with the quantitative result of no significant changes in enjoyment (M gain = -.004, SD = 1.013, p = .981), and indicated that while drawing was enjoyable, it simply was not enjoyable enough to elicit a significant increase in enjoyment from baseline camping. The confidence participants felt from creating a visual that was tangible was resounding with participants who enjoyed visual expression through drawing theirs.

In the control group, motivation was varied. Some people felt inspired while others reported neutral or disinterest. Some positives included participants explaining, “Yes, I guess, because writing with the storyboard made me feel like I’m on the right way,” explaining how drawing support organized ideas. Other participants were less positive or simply uninterested with comments like, “No because I don’t love writing.” In terms of variability in feelings, this programming was equivalent to the drop of motivation in the phase (M gain = -.345, SD = .919, p = .031, Cohen’s d = -.375); drawing was exciting to some and thank goodness as being passive were reflected in final passive.

Time and drawing skills in the control group were also mentioned to be challenges. Time was noted by participants, “We were short on time so I couldn't finish it completely,” while other participants mentioned struggles of drawing ability, “No, it didn’t help as I am not good at drawing so I get bored while drawing.” As such, it is possible that time was shorted in the motivation work and drawing did not anticipate interest.

In terms of time, this reasoning is responded to with moderate support from the appreciation phase originally motivation work (Cohen’s d = -.375) parity respond, (M gain = .021, SD = .768, p = .868), can be conveyed. For participants less skilled or add confidence for drawing we found the task rather a frustration not a motivation value. Drawing’s impact on organizing thoughts

for writing varied by individual. Some found it helpful, noting, "Sometimes it helps me to organize my thoughts it is a better way to figure out what I think," while others struggled, as in, "No, because I don't have the talent to draw. And this challenges me." This mixed effectiveness explains the lack of significant shifts in enjoyment or anxiety, as drawing's benefits were not universal and depended heavily on participants' skills and preferences.

In conclusion, the qualitative analysis reveals that both SORA and drawing offered distinct benefits. Both of them appealed to different facets of the writing process, which explains the lack of significant quantitative differences.

Integration of Findings

In the current study, a mixed-methods approach was implemented in order to investigate the affective outcomes of AI-generated storyboards. The quantitative findings indicated statistically modest differences between groups, with the most interesting finding being a noted motivational drop in the control group. The qualitative findings expanded the descriptions of the quantitative results. Students in the experimental group noted high levels of engagement, fun, and creative support - attributes which, likely shaped motivational stability.

The triangulation of the findings suggests that merely relying on quantitative measures may under-represent potential affective benefits of AI tools such as SORA. The complementary evidence suggests, while shifts in quantitative data were limited by sample size and length of the intervention, student perspectives captured in interviews provided strong evidence for the motivational affordances of AI-mediated storyboarding. Taken together, the findings highlight the importance of incorporating student voice and engagement measures in evaluating affective outcomes within technology-mediated language learning.

Discussion

The present study investigated the affective outcomes of using AI-generated video storyboards via SORA in Content and Language Integrated Learning (CLIL) writing instruction, examining anxiety, motivation, enjoyment, and engagement. A key finding was a significant decline in motivation in the control group, contrasted by the maintenance of motivation in the experimental group using SORA. This result aligns with emerging evidence that AI-enhanced instruction can stabilize learner affect in language learning contexts, particularly by reducing writing-related anxiety and boosting engagement through scaffolded, multimodal support. The motivational sustenance that the SORA group showed in contrast to the drop in the control group significantly indicates the facilitation provided by AI to affective stability. Printer (2023) puts forward a very similar end through Self-Determination Theory, claiming that its narrative-based teaching is effective in fulfilling the needs for autonomy and competence among learners. In this study, novelty and interactivity of SORA must have provided that sense of autonomy with little performance pressure, hence the absence of drop in motivation like that observed among the control group participants. While motivation diverged, anxiety levels remained more or less the same between groups, which yields disparate findings in the literature. Mettewie et al. (2024), in an extensive longitudinal CLIL study, pointed out, initial buoyancy and happiness decline after a while. This can reinforce the findings from the current study, hinting that SORA assisted in motivating but perhaps lessened anxiety only initially or even over longer exposure

times. However, the meta analysis of Wu (2024) shows evidence of AI's greater potential to affect social-emotional endpoints such as anxiety and motivation, which means that more intense time plus use of SORA could produce impressive benefits in the emotional spectrum.

The SORA group indicated high engagement, a finding in line with research on AI-assisted storytelling. Chen (2024) found that AI-assisted digital storytelling improved affective experiences through visual and auditory elements, with all creativity lying with the human. Likewise, Cheung and Shi (2025) found that students who co-created a story with an AI companion retained agency in the use of that story, as it was enhancing the human creative storytelling process rather than supplanting, or usurping the human role or space. Participants did note issues with SORA, namely prompt specificity and time. Woo et al. (2023) also found that AI can push students' creative edges; however, prompt issues were discussed in the same way, reducing the overall effects of the outcome. These practical challenges may have mitigated SORA's overall impact. The evidence of the literature suggests that positive effects on writing instruction depend on emotional dynamics, particularly anxiety and creativity. Indeed, Wang et al. (2021), found that state anxiety was negatively related to performance and creativity in EFL writing development, thus they recommend to find a way to index emotional regulation and to use emotional regulation tools (AI scaffolding) in writing. Anxiety remained constant throughout the study, indicating a probable motivational effect of SORA that may have lessened performance-related anxiety as it pertains to creativity. These echoes have been seen in Ramezanzadeh and Ebadi's (2024) investigation of AI-assisted emotionally driven storytelling, or how supported vulnerability can encourage language learning in different contexts. Although SORA was used as a mechanism for learner engagement and creativity, it falls under the umbrella of a much larger phenomenon: digital multimodal storytelling. Isik and Isik (2021) found that learner-generated stories were much more effective than pre-made stories in building peer connection and deeper language ability, as did Davy Tsz Kit et al. (2022) who also found AI story writing enhanced story competency and AI literacy in primary aged students. Similarly, SORA's video storyboards improved the writing experience through a more classroom-centered, multimodal approach. However, SORA simply provides the illusion of deeper writing, as Stewart and Zheng (2024), warned against reliance on AI-generated processes as often these outputs border on superficiality, with no depth or context.

There are also many limitations to the current research investigating the use AI-generated video storyboards in CLIL for pedagogical writing instruction. While the size for each group (experimental and control) was 36 participants the findings may not extend to wider populations. A larger, more heterogeneous sample would have improved statistical power and applicability. In addition, the intervention was of short duration, and thus it was impossible to assess any long-term effects; it is not possible to conclude whether gains like sustained motivation would persist. The author also utilized self-reported measures to assess affective outcomes (anxiety, motivation, enjoyment, and engagement) which may have also introduce biases—specifically the possibility participants were socially desirable in their responses; in turn, some of the conclusions regarding self-reported responses could have been validated by more objective measures (eg. writing performance) to bolster and strengthen the findings in a robust way. In addition, the study made no attempt to account for individual differences

creativity, writing ability, or access to AI tools before the intervention; any of these could have had an impact on the overall results.

Finally, study findings are limited to a CLIL context and will only be relevant to educational contexts outside of CLIL and different areas of study. Although quantitative findings were not statistically significant across most metrics, many effect sizes (e.g., Cohen's d = −.465 for motivation) suggest minor practical meaning. The process of future studies confirming those trends should include surveying larger samples with planned instruction and educational usage. Qualitative responses provided explanatory depth for the motivational maintenance observed in the SORA group. For example, while quantitative findings indicated no significant gains in enjoyment (use); students described high levels of engagement, enjoyment, and visualization support. This complementarity suggests affective changes may be more likely captured in qualitative findings in short-term interventions.

Conclusion

This study examined the affective outcomes of anxiety, motivation, enjoyment, and engagement of using AI-designed video storyboards (with SORA) to support writing within a CLIL context. It showed the SORA group maintained motivation throughout the intervention period while the control group maintained significant declines. However, no significant differences were found between groups with respect to anxiety, enjoyment or engagement. Qualitative data demonstrated the novelty of SORA and interactivity as the main motivators for maintaining motivation over time, but it was also influenced by external barriers like generating accurate prompts and time. This suggests that AI tools like SORA can help maintain motivation during language learning situations though did not provide an impact on other affective outcomes and there will be barriers to installing to be fully utilized. This study contributes to a growing literature on integrating technology into the classroom to support student well-being by showcasing that while AI can be motivating, there remain improvements and design aspects to consider. To build on the findings of this study, future studies should further address the limitations of the study as well as consider other possibilities of AI-based methods of development. First, researchers should endeavor to recruit larger and more heterogeneous samples to improve the generalizability of findings, as well as detect some of the finer-grained effects that may have been masked by the smaller sample of this study. Second, the value of self-reports could be enhanced still by looking for objective measures, such as quantitative means of assessing writing quality as well as measuring more physiological items (e.g., heart rate variability) to chart anxiety. Third, longitudinal studies are called to learn about whether SORA can retain its motivational nature over time, and if anxiety or enjoyment shifts over time. Fourth, it would plainly show how individual differences such as creativity, writing ability, and experience with AI tools contribute to the outcome of the intervention, which would lead to further tailoring of the experience. Fifth, inquiries into other non-CLIL contexts and across subject areas will serve to substantiate the transferable application of AI generated storyboards more broadly across the educational milieu. Finally, prompt templates or real time feedback prompts developed could serve to enhance the experience (if any), the usage, effectiveness, and access of the tool within teacher education programs.

The results of this research provide applicable recommendations to educational stakeholders,

outlining concrete strategies for implementing AI tools such as SORA to great effect. A huge motivational factor for students when using SORA in CLIL writing instruction is that it is fun and interactive. Teachers should run pre-intervention training to instruct students about writing effective prompts, demonstrating appropriate phrasing through examples that depict expected visual output. Thus, allowing for 60 to 90 minutes for students to explore SORA for the first time means that they can get through the learning curve and get the most out of the tool. Teachers should assign reflective tasks that encourage critical engagement and help affirm student agency and creative ownership. Examples include comparing AI-generated storyboards with their original vision. To do this, EFL teachers can embed a brief "prompt-engineering clinic" into the pre-task phase, using think-aloud modelling and scaffolded worksheets that guide learners from teacher-provided ones toward fully independent prompt creation. The study indicates to researchers that there is a need to investigate the longer-term affective benefits of AI even further. Researchers may design their studies to longitudinally track motivation and engagement across semesters, mixing data collection methods (e.g., surveys with writing sample analyses) to capture both subjective and objective effects. By recognizing how learners' characteristics, such as their previous writing ability, or preferences for technology, could have an impact, there is the potential for personalization of the AI intervention. Because of the engaging experience offered by the video storyboards, instructional designers can also see how multimodal design features can be additive to the educational environment. For example, instructional designers might add contextually aware scaffolds such as visual metaphors that fulfill a particular lesson objective or interactive step-by-step tutorials, with their permutations, and even analytics dashboards that would allow the teacher to monitor cognitive load and engagement in real-time. They also need to improve the AI hubs with such features as automatically generated prompts that respond to user output, or by providing guided tutorials to assist students in developing their storyboards. The goal of these improvements is to reduce usability challenges while aligning more closely to real classroom practice. At the education policy level, initiatives need to be made to develop specific guidelines on data privacy, fund development grant programs for micro-credentials for teachers to certify them in the ethical use of AI, and to conduct pilot studies before wide implementation to guarantee the ethical and effective use of AI for writing instruction. Implementing these initiatives would help key stakeholders to realize the promise of the potential of AI to improve writing instruction and improve the practical challenges that were identified through this study.

References

Ardasheva, Y., Tong, S. S., & Tretter, T. R. (2012). Validating the English Language Learner Motivation Scale (ELLMS*). Learning and Individual Differences, 22*(4), 473–483 https://doi.org/10.1016/j.lindif.2012.03.001

Behrens, K. A., Marbach-Ad, G., & Kocher, T. D. (2024). AI in the genetics classroom: A useful tool but not a replacement for creative writing. *Journal of Science Education and Technology.* https://doi.org/10.1007/s10956-024-10160-6

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Chan, C. K. Y., & Tsi, L. H. Y. (2023). The AI revolution in education: Will AI replace or assist teachers in higher education? *arxiv.* https://doi.org/10.48550/arXiv.2305.01185

Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access, 8*, 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

Chen, Y.C. (2024). Perceptions of AI-facilitated creativity in language education: A study on digital storytelling. *The JALT CALL Journal, 20*(3), xx–xx. https://doi.org/10.29140/jaltcall.v20n3.2089

Cheng, Y. S. (2004). A measure of second language writing anxiety. *Journal of Second Language Writing, 1*3(4), 313–335. https://doi.org/10.1016/j.jslw.2004.07.001

Cheung, L., & Shi, H. (2025). Co-creating stories with generative AI. *Australian Review of Applied Linguistics.* https://doi.org/10.1075/aral.24101.che

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Lawrence Erlbaum.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated learning*. Cambridge University Press.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches (4th ed.)*. Sage.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign-language classroom. *Studies in Second Language Learning and Teaching, 4*(2), 237–274. https://doi.org/10.14746/ssllt.2014.4.2.5

Field, A. (2013). *Discovering statistics using IBM SPSS statistics (4th ed.)*. Sage.

Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education (6th ed.)*. McGraw-Hill.

Gardner, R. C. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. Edward Arnold.

Gordon, B. M. (2011). *Artificial intelligence: Approaches, tools and applications*. Nova Science Publishers.

Hayes, J. R. (2012). Modeling and remodeling writing. *Written Communication, 29*(3), 369–388. https://doi.org/10.1177/0741088312451260

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning.* Center for Curriculum Redesign.

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, research issues and applications of artificial intelligence in language education. *Educational Technology & Society, 26(1), 112–131*. https://doi.org/10.30191/ETS.202301_26(1).0009

Isik, A., & Isik, B. (2021). We are reading the stories we write. *Electronic Journal of Foreign Language Teaching*, 18. https://doi.org/10.56040/isik1812

Ivanytska, N., Koliasa, O., Kovalevska, T., Matsera, O., & Tkachuk, T. (2024). Analyzing the possibilities of implementation of AI and social networks in teaching foreign-language students: Ukrainian universities case study. *Arab World English Journal (Special Issue on ChatGPT), 306*–318. https://doi.org/10.24093/awej/ChatGPT.21

Krashen, S. D. (1982*). Principles and practice in second language acquisition*. Pergamon Press.

Lee, H., & Lee, J. H. (2024). The effects of AI-guided individualized language learning: A meta-analysis. *Language Learning & Technology, 28*(2), 134–162. https://hdl.handle.net/10125/73575

Li, C., Zhang, L. J., & Jiang, G. Y. (2021). Conceptualisation and measurement of foreign-language learning burnout among Chinese EFL students. *Journal of Multilingual and Multicultural Development, 45(*4), 906–922. https://doi.org/10.1080/01434632.2021.1931246

López-Medina, E. F., & Casado, J. M. (2024). CLIL, a competence-based coping strategy against foreign-language anxiety in higher education. *Discover Psychology, 4,* 68. https://doi.org/10.1007/s44202-024-00187-6

MacIntyre, P. D., & Gardner, R. C. (1991). Investigating language-class anxiety using the focused-essay technique. *The Modern Language Journal,* 75(3), 296–304. https://doi.org/10.1111/j.1540-4781.1991.tb05358.x

McCarthy, J., Minsky, M. L., Rochester, N., & Shannon, C. E. (1955). A proposal for the Dartmouth summer research project on artificial intelligence. *AI Magazine, 27*(4), 12–14.

Mettewie, L., Van Mensel, L., De Smet, A., & Galand, B. (2024). A longitudinal test of the impact of CLIL on language emotions and learning motivation. *System, 117*, 103355. https://doi.org/10.1016/j.system.2024.103355

Mohammed, S. J., & Khalid, M. W. (2025). Under the world of AI-generated feedback on writing: Mirroring motivation, foreign language peace of mind, trait emotional intelligence, and writing development. *Language Testing in Asia, 15*(7), Article 43. https://doi.org/10.1186/s40468-025-00343-2

Niloy, A. C., Akter, S., Sultana, N., Sultana, J., & Rahman, S. I. U. (2024). Is ChatGPT a menace for creative-writing ability? *An experiment. Journal of Computer Assisted Learning, 40*(2), 919–933. https://doi.org/10.1111/jcal.12929

Noels, K. A. (2001). Learning Spanish as a second language: Learners' orientations and perceptions of their instructors' communication style. *Language Learning, 51*(1), 107–144.

O'Brien, H. L., Cairns, P., & Hall, M. (2018). A practical approach to measuring user engagement. *International Journal of Human–Computer Studies, 112*, 28–39. https://doi.org/10.1016/j.ijhcs.2018.01.004

Ouyang, F., & Jiao, P. (2021). Artificial-intelligence in education: The three paradigms. Computers and Education: *Artificial Intelligence, 2*, 100020. https://doi.org/10.1016/j.caeai.2021.100020

Paivio, A. (1991). Dual-coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287. https://doi.org/10.1037/h0084295

Printer, L. (2023). Positive emotions and intrinsic motivation: A self-determination theory perspective on using co-created stories in the language-acquisition classroom. *Language Teaching Research. Advance online publication.* https://doi.org/10.1177/13621688231204443

Ramezanzadeh, A., & Ebadi, S. (2024). AI-generated storytelling for socially just EFL pedagogy: Exploring learner emotions through Fraser's justice theory. *System, 116,* 103588. https://doi.org/10.1016/j.system.2024.103588

Roll, I., & Wylie, R. (2016). Evolution and revolution in artificial intelligence in education. *International Journal of Artificial Intelligence in Education, 26*(2), 582–599. https://doi.org/10.1007/s40593-016-0110-3

Song, C., & Song, Y. (2023). Enhancing academic-writing skills and motivation: Assessing the efficacy of ChatGPT in AI-assisted language learning for EFL students. *Frontiers in Psychology, 14*, Article 1260843. https://doi.org/10.3389/fpsyg.2023.1260843

Stewart, N., & Zheng, Y. (2024). Generative AI's recolonization of EFL classrooms. Australian Review of Applied Linguistics. *Advance online publication.* https://doi.org/10.1075/aral.24091.ste

Sweller, J. (2010). *Cognitive load theory: Recent theoretical advances.* In J. L. Plass, R. Moreno, & R. Brünken (Eds.), Cognitive load theory (pp. 29–47). Cambridge University Press.

Tsz Kit, N. G. D., Luo, W., Chan, H., & Chu, S. (2022). An examination of primary students' development in AI literacy through digital-story writing. *Computers and Education: Artificial Intelligence, 3,* 100054. https://doi.org/10.1016/j.caeai.2022.100054

Verdenhofa, O., Kinderis, R., & Berjozkina, G. (2024). AI as a creative partner: How artificial intelligence impacts student creativity and innovation. *Baltic Journal of Economic Studies, 10*(5), 111–119. https://doi.org/10.30525/2256-0742/2024-10-5-111-119

Villabona, N., & Cenoz, J. (2022). The integration of content and language in CLIL: A challenge for content-driven and language-driven teachers. *Language, Culture & Curriculum, 35*(1), 36–50. https://doi.org/10.1080/07908318.2021.1910703

Wang, D. (2024). Teacher- versus AI-generated (Poe application) corrective feedback and language learners' writing anxiety, complexity, fluency, and accuracy. *International Review of Research in Open and Distributed Learning, 25*(3), Article 7646. https://doi.org/10.19173/irrodl.v25i3.7646

Wang, H., Fu, T., Du, Y., Gao, W., Huang, K., Liu, Z., … Marks, D. (2023). Scientific discovery in the age of artificial intelligence. *Nature, 620*(7972), 47–56.

https://doi.org/10.1038/s41586-023-06221-2

Wang, H.-C. (2021). Exploring the relationships of achievement motivation and state anxiety to creative-writing performance in English as a foreign language. *Thinking Skills and Creativity, 42,* 100948. https://doi.org/10.1016/j.tsc.2021.100948

Woo, D. J., Guo, K., & Salas-Pilco, S. Z. (2024). Writing creative stories with AI: Learning designs for secondary-school students. *Innovation in Language Learning and Teaching.* https://doi.org/10.1080/17501229.2024.2384884

Woo, D. J., Guo, K., & Susanto, H. (2023). Exploring EFL students' prompt engineering in human-AI story writing: An Activity-Theory perspective. *arXiv.* https://doi.org/10.48550/arXiv.2306.01798

Wu, X. Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 116*, 103498. https://doi.org/10.1016/j.system.2024.103498

Zhu, X., Zhan, J., & Yao, Y. (2022). A latent-profile analysis of Chinese EFL learners' enjoyment and anxiety in reading and writing: Associations with imaginative capacity and story-continuation writing performance. *International Review of Applied Linguistics in Language Teaching. Advance online publication.* https://doi.org/10.1515/iral-2022-0105

Biodata

Esen Genç was born on December 5, 1984. She received education at Ondokuz Mayıs University in the Department of English Language Teaching. She graduated from university in 2007. She got a Master's Degree in the same department at the same university in 2014. She got her Doctoral Degree in the Ph.D. Program on English Language and Literature at İstanbul Aydin University in 2021. She is working as an Assist. Prof. Dr. at the Department of Translation and Interpreting at Ondokuz Mayıs University, and she carries out her academic studies in the same field.