

Increasing students' engagement in English lessons through video lectures in a primary school ¹⁰⁸⁰

Nguyen Thi Thao

nguyenthithao27031999@gmail.com
Thuy Duong Primary School, Hai Phong city, Viet Nam

Abstract

This action research study investigates how video lectures enhance engagement and transform learning attitudes among 42 Grade 4 EFL students at Thuy Duong Primary School. Conducted over 12 weeks, the research compared traditional linear YouTube cartoons in Cycle 1 with interactive, AI-generated video lectures featuring local school contexts and virtual teacher avatars in Cycle 2. Data collection was triangulated using classroom observation checklists, student questionnaires, and a teacher's reflective journal. Quantitative analysis revealed a statistically significant improvement in student engagement across behavioral, emotional, and cognitive dimensions, alongside a positive shift in learning attitudes. Qualitative findings confirmed that interactive AI features and pause-and-predict techniques effectively lowered the learners' affective filters, transitioning them from passive consumers to active speakers. The study implies that teachers should transition from passive consumers of generic online content to active designers of customized, locally relevant digital resources. Utilizing AI to personalize learning environments drastically minimizes extraneous cognitive load and cultural barriers for young learners. Furthermore, embedding structured interactive pauses transforms standard viewing into generative processing, while virtual avatars provide a psychologically safe scaffolding space that builds communicative confidence before students engage in real-world interactions.

Keywords: Student engagement; Video lectures; Primary school; Action research; Generative AI

1. Introduction

Despite robust macroscopic policy initiatives such as the National Foreign Language Project 2020, translating institutional goals into public primary school classrooms in Vietnam faces significant practical obstacles. In regional, public primary settings, particularly within suburban or rural locales like Thuy Nguyen Ward, transitioning from traditional, teacher-centered routines to learner-centered Communicative Language Teaching (CLT) remains a major pedagogical hurdle. Instructors frequently contend with large class sizes, restricted instructional hours, and a heavy systemic dependence on textbooks. Consequently, Grade 4 English lessons often revert to rote learning practices that prioritize the mechanical memorization of vocabulary and grammatical structures over real-time communication. This traditional instructional environment fails to stimulate the natural curiosity of young children, resulting in an unexciting learning atmosphere characterized by low student engagement and a widespread hesitancy to speak.

Grade 4 represents a critical and sensitive developmental demographic that justifies targeted empirical intervention. While Grade 3 functions as an introductory year where high student interest is sustained by the absolute novelty of the subject, this initial situational motivation frequently diminishes by Grade 4. Curricular demands naturally accelerate at this stage, introducing more complex sentence structures and an expanded lexicon that require greater cognitive effort. Psychologically, children aged 9 to 10 are navigating a transitional phase from play-based learning to formal academic instruction demanding sustained attention and discipline. As active meaning-makers, young learners depend heavily on continuous sensory stimulation; if classroom tasks lack dynamism, their interest rapidly declines. Prolonged disengagement at this stage is a

severe issue, as early negative experiences with language acquisition frequently induce long-term language anxiety, passivity, and a permanent deficit in communicative confidence.

To address this polarization, educational research highlights the necessity of fostering multi-dimensional student engagement, which serves as the primary engine for academic success and competence development. Moving beyond surface-level behavioral compliance (such as sitting quietly), contemporary frameworks conceptualize engagement as a tripartite construct comprising behavioral (participation and effort), emotional (interest and confidence), and cognitive (metacognitive investment and deep understanding) dimensions. Because these three dimensions operate in a dynamic, reciprocal motivational loop where emotional interest triggers behavioral action, which in turn supports cognitive processing, instructional methods must target all three components simultaneously.

Technology-assisted instruction, specifically the integration of video lectures, presents a prominent solution to this challenge. Grounded in the Cognitive Theory of Multimedia Learning (CTML), well-designed videos utilize the human brain's dual channels to process auditory and visual information in parallel, thereby optimizing working memory capacity and facilitating superior comprehension. While international literature extensively underscores the value of instructional videos for capturing learner attention, a distinct gap remains within the Vietnamese primary English as a Foreign Language context. Existing local studies predominantly focus on summative learning outcomes (such as test scores) or generalized student attitudes, leaving a critical gap regarding how specific video design parameters shape behavioral, emotional, and cognitive engagement in real time.

Furthermore, conventional authentic videos often carry a heavy cultural burden, creating a cognitive disconnect for young beginners in localized environments. To bridge these empirical and practical gaps, this action research introduces an innovative instructional paradigm by leveraging generative artificial intelligence (AI) tools (such as Google Veo and HeyGen) to construct personalized, interactive video materials that reflect the students' actual school environment and embed virtual speaking partners. Driven by the immediate practical need to dismantle classroom passivity and lower the learners' affective filters, this study provides a systematic, triangulated evaluation of how AI-driven contextualization and interactive techniques can successfully transition young learners from passive video consumption to proactive, confident oral communication.

2. Literature Review

2.1. *Student engagement and learner attitudes in primary schools*

Student engagement is a multi-dimensional construct comprising behavioral (active participation), emotional (enthusiasm and interest), and cognitive (deep psychological investment) dimensions. In primary English as a Foreign Language education, fostering this engagement is crucial because young learners are active meaning-makers who depend heavily on sensory stimulation and a supportive environment to maintain focus and take communicative risks. However, local Vietnamese public primary classrooms face structural challenges, such as large class sizes, limited instructional time, and an over-reliance on traditional textbooks, that often cause teachers to revert to rote grammar memorization. This environment inadvertently induces student passivity and speaking anxiety. Dismantling this passivity requires a shift in learner attitude, which consists of cognitive beliefs regarding utility, affective feelings of enjoyment, and conative/behavioral intentions to interact (Baker, 1992; Eagly & Chaiken, 1993). Video lectures can successfully optimize this attitudinal structure by simultaneously clarifying meanings, reducing classroom boredom, and generating a higher behavioral readiness to communicate.

2.2. Theoretical foundations and design of video lectures

The integration of video lectures into primary EFL classrooms is theoretically supported by Mayer's (2009) Cognitive Theory of Multimedia Learning, which operates on three core assumptions: dual-channel processing (auditory and visual streams are processed separately), limited capacity (working memory can only hold a restricted amount of information at once), and active processing (learning requires actively selecting, organizing, and integrating new inputs with prior knowledge). To optimize these mental pathways and prevent split-attention effects or cognitive overload that elevates learner anxiety, educational media must adhere to three central design principles outline by Ibrahim et al. (2012): segmenting content into user-paced chunks, signaling key language items with explicit visual or auditory cues, and weeding out irrelevant background noise or decorative animations. When tailored for young EFL learners, video design generally utilizes instructional animation to lower extraneous load and optimize vocabulary retention, teacher-recorded videos or virtual AI teacher avatars to create an instructor presence that lowers the affective filter, and interactive mechanism features like pause-and-predict checkpoints to securely anchor student attention against distraction.

2.3. The impact of video lectures on engagement dimensions

Integrating well-designed video lectures serves as a powerful driver to enhance multi-dimensional student engagement. Behaviorally, dynamic moving pictures counteract attention decay, acting as cognitive hooks within a micro-learning threshold (under three minutes) that accommodates the short attention spans of young learners. This captures attentional focus and triggers voluntary hand-raising and active tracking. Emotionally, video lectures serve as a critical buffer that lowers foreign language performance anxiety by replacing high-pressure direct questioning with attractive, supportive animated narratives. Cognitively, dual-coded media combines visual pictures with spoken sounds to provide a structural support system for the brain. This visual scaffolding enables primary students to regulate their learning pace, internalize vocabulary, replicate native pronunciation, and transition from broken words to complete, grammatically correct sentences without relying on surface-level rote memorization.

2.4. Previous literature synthesis and research gap

Ample international and domestic literature highlights the efficacy of video-based language teaching. Internationally, animated videos drastically accelerate vocabulary retention and listening comprehension among young beginners. For instance, Minalla (2024) reported vocabulary mean score jumps from 11.84 to 22.18 when using animations. In Vietnam, studies confirm that commercial YouTube cartoons lower student anxiety and improve listening and speaking proficiency scores (Le & Tran, 2020), while short video structures maximize learner satisfaction when explicit teacher presence is maintained (Nguyen & Pham, 2025). However, three critical research gaps remain:

Demographic and Environmental gap: Seminal multimedia video design studies focus heavily on adult populations in higher education or MOOCs, leaving public primary school classrooms with large class sizes under-researched.

Cultural and Contextual gap: Most research relies on pre-existing Western cartoons, which present a cultural disconnect that overloads young learners' processing channels (Guariento & Morley, 2001). There is a lack of research examining video materials customized via Generative AI to reflect local culture and familiar Vietnamese school settings.

Methodological and Analytical gap: Local primary studies focus almost exclusively on pre-and-post learning outcomes, failing to track the fine-grained behavioral, emotional, and cognitive engagement process sequentially in a typical public school setting.

In summary, the literature review establishes that student engagement is a multi-dimensional construct comprising behavioral, emotional, and cognitive layers, all of which are critical for developing communicative competence in young EFL learners. While standard, textbook-heavy methodologies in Vietnamese primary schools frequently lead to classroom passivity and performance anxiety, multimedia instruction guided by Mayer's Cognitive Theory of Multimedia Learning offers a scientifically proven framework to reduce cognitive overload and capture attention. However, existing research reveals a critical gap: standard educational materials often introduce a cultural disconnect for regional beginners, and local studies rarely evaluate the fine-grained, concurrent process of behavioral, emotional, and cognitive engagement in public school settings. This provides the empirical rationale for the present action research, which utilizes personalized, AI-contextualized interactive video lectures to bridge the gap between abstract

3. Methodology

This study employed a two-cycle participatory action research design based on the iterative spiral model by Kemmis and McTaggart (2000), aimed at resolving an ongoing problem of engagement polarization and language anxiety in a real classroom setting. The primary research objectives were to examine the longitudinal impact of video lectures on student engagement and to explore learners' attitudes toward shifting from linear media to interactive, AI-contextualized video formats. To evaluate the intervention, the study addressed two central research questions focusing on how video lectures improve multi-dimensional engagement and what explicit attitudes students hold toward this instruction. Utilizing purposive sampling, the research sample consisted of all 42 Grade 4 students (22 females and 20 males, aged 9–10) in Class 4A1 at Thuy Duong Primary School in Hai Phong City. Data collection instruments were strictly triangulated to capture both external behaviors and internal perceptions, comprising a structured 4-point classroom observation checklist adapted from Hoang (2023) and Skinner and Pitzer (2012), two bilingual post-intervention questionnaires utilizing a child-friendly 4-point visual emoji scale (Read, 2008), and a teacher's reflective journal based on Richards and Lockhart's (1994) reflective teaching framework. The 12-week research procedure was organized into four distinct phases aligning with the Global Success 4 curriculum: a preparation phase to validate instruments and pilot tools; Cycle 1 (Weeks 1–6) utilizing standard linear YouTube animations; a two-week reflection and AI material production phase (Weeks 7–8); and Cycle 2 (Weeks 9–12) introducing interactive “pause-and-predict” techniques embedded within customized AI-generated video lectures (via Google Veo and HeyGen) featuring the students' actual school surroundings. Quantitative data analysis was conducted via SPSS (version 26.0) using descriptive statistics and parametric paired samples t-tests to measure significance across timelines, while qualitative notes from the teacher's journals were processed through Braun and Clarke's (2006) six-step thematic analysis model to provide deep contextual insights.

4. Results and discussion

4.1. Findings from the questionnaire results

The quantitative findings derived from the classroom observation checklists and post-intervention student surveys provide empirical evidence regarding the impact of video lectures on the classroom ecosystem. The data systematically traces changes across two distinct action research phases: Cycle 1 (traditional linear video playback) and Cycle 2 (interactive, AI-generated contextualized video delivery).

4.1.1. Analysis of Student Engagement

To answer the first research question regarding the extent to which video lectures improve Grade 4 student engagement, the researcher monitored the entire class (N=42) across three core dimensions utilizing a 4-point tracking scale ranging from 0 (Not Observed) to 3 (Consistently). The descriptive statistics and the growth metrics between the two cycles are detailed in Table 4.1 below:

Table 4.1. Descriptive Statistics of Student Engagement (N=42)

Engagement Dimension	Cycle 1 Mean	Cycle 2 Mean	Growth (M2-M1)	Interpretation (Cycle 2)
Behavioral	1.52	2.12	+0.60	High Engagement
Emotional	1.60	2.25	+0.65	High Engagement
Cognitive	1.45	2.05	+0.60	High Engagement
Overall	1.52	2.14	+0.62	High Engagement

Note: Scale interpretation: 0.00–1.00 (Low); 1.01–2.00 (Moderate); 2.01–3.00 (High).

As illustrated in Table 4.1, the total mean score for the class experienced a clear upward shift, rising from a moderate baseline of 1.52 in Cycle 1 to a high engagement level of 2.14 in Cycle 2. This longitudinal growth highlights specific dimensional shifts within the classroom:

- Behavioral Dimension (1.52 - 2.12): In Cycle 1, student behavior was primarily characterized by passive screen-watching. In Cycle 2, the implementation of interactive checkpoints successfully converted passive attention into active visible outputs, such as voluntary hand-raising, physical tracking, and audible responses.
- Emotional Dimension (1.60 - 2.25): This sector yielded the largest single margin of growth (+0.65). Triangulated data indicates that incorporating localized school visuals via AI personalization drastically reduced language anxiety and prompted evident affective reactions, such as smiling and mutual excitement.
- Cognitive Dimension (1.45 - 2.05): Cognitive metrics grew steadily (+0.60), verifying that the step-by-step “pause-and-predict” framework compelled students to deploy deep thinking and actively generate contextual linguistic hypotheses instead of relying on mechanical, surface-level memorization.

To confirm that these observable shifts were mathematically rigorous, a parametric paired samples *t*-test was executed on the collected observation logs. The results showed a statistically significant difference between the two cycles ($t = 2.65$, $p = 0.012$), demonstrating that the multidimensional engagement spike was a direct consequence of the interactive AI video architecture and not a product of random variance ($p < 0.05$).

4.1.2. Analysis of Student Attitudes

The second research question evaluated the explicit self-reported perceptions of the learners (N=42). Core longitudinal and cycle-specific survey prompts were scored on a child-friendly 4-point visual emoji metric where 1 represented “very unhappy” and 4 represented “very happy”.

A parametric paired samples *t*-test conducted on the core longitudinal survey prompts (Q1-Q7) confirmed a statistically significant positive trajectory in learner attitudes ($t = 2.82$, $p = 0.008$). The baseline class mean moved from a neutral 2.65 (SD = 0.51) in Cycle 1 into a highly positive 3.05 (SD = 0.46) threshold during Cycle 2. The contraction of the standard deviation indicates an increased consensus of positive attitudes across the diverse proficiency spectrum.

To provide a detailed item-by-item overview, Table 4.2 presents the mean scores for all ten questionnaire statements across both intervention cycles:

Table 4.2. Comparison of Mean Scores by Questionnaire Item

No.	Statement (Simplified for Grade 4)	Cycle 1 Mean	Cycle 2 Mean	Difference	Cluster Category
Q1	Do you think English lessons with videos are fun?	2.60	3.00	+0.40	Enjoyment & Interest
Q2	Do videos help you focus better on the teacher's lesson?	2.70	3.05	+0.35	Attention & Focus
Q3	Do videos help you understand and remember words more easily?	2.65	2.95	+0.30	Understanding & Memory
Q4	Do you feel happy and relaxed when learning with videos?	2.45	2.90	+0.45	Confidence / Anxiety Alleviation
Q5	Do videos make you want to talk and interact more with your friends?	2.60	2.95	+0.35	Classroom Interaction
Q6	Do you find the video stories interesting?	2.75	3.10	+0.35	Enjoyment & Interest
Q7	Can you finish the class tasks better when there are videos?	2.60	2.95	+0.35	Task Completion
Q8	<i>Cycle-Specific Context Question*</i>	2.75	3.25	+0.50	AI Feature Personalization
Q9	<i>Cycle-Specific Audio/Music Question*</i>	2.60	3.15	+0.55	AI Feature Personalization
Q10	Do you want to watch more videos like these in class?	2.85	3.20	+0.35	Enjoyment & Interest

**Note: Items 8 and 9 contained cycle-specific prompts: Cycle 1 evaluated traditional cartoon visuals and music; Cycle 2 evaluated the AI-generated school context and the helpfulness of the AI Avatar.*

The detailed item analysis revealed key psychological breakthroughs within three primary clusters:

- Enjoyment and Interest (Q1, Q6, Q10): General situational enjoyment advanced systematically. Item 1 climbed from 2.60 to 3.00. While children structurally welcomed traditional animated entertainment in Cycle 1, the interactive components introduced in Cycle 2 rendered the academic tasks intrinsically motivating because they were actively participating in predicting the narrative's progression rather than sitting as detached onlookers.
- Confidence and Anxiety Alleviation (Q4, Q5): Reflecting a vital psychological milestone, Item 4 ("Do you feel confident to speak?") experienced a substantial jump of +0.45, climbing from a low 2.45 to 2.90. The data implies that the non-judgmental presence of the digital AI Avatar established a psychologically safe training zone, allowing anxious or low-proficiency students to voice responses without fear of peer embarrassment.
- AI Feature Personalization (Q8, Q9): Cycle 2 specific variables achieved the highest absolute ratings in the entire profile. Item 8 ("Excitement about school context") peaked at 3.25, validating that seeing their real-world school surroundings explicitly embedded inside the digital media strongly galvanized student focus. Concurrently, Item 9 ("Helpfulness of AI Avatar") reached a mean of 3.15, proving that a virtual speaking partner acts as an effective scaffold to drive primary language interactions.

4.2. Findings from the Reflective Journal

While quantitative metrics track broad classroom trends, qualitative data compiled from the teacher's reflective journal provides an in-depth understanding of the underlying causes behind these numbers. Systematically analyzed through Braun and Clarke's (2006) six-step thematic framework, the journal entries reveal three prominent pedagogical themes that shaped student participation.

4.2.1. Theme 1: Engagement dynamics

The qualitative logs indicate a fundamental shift not just in the volume of engagement, but in its absolute quality between the two implementation phases.

- Cycle 1-High Attentional Capture, Low Verbal Production: During the first six weeks, standard linear YouTube animations were highly successful at securing basic attention. The teacher noted that students maintained “100% eye contact with the visual screen” and displayed high physical energy as soon as the background theme music commenced. However, this interaction remained entirely receptive. The journal entries documented a clear “output barrier,” noting that students treated educational video content exactly like a commercial television broadcast designed for passive consumption. When the video was paused for direct questioning, the comfortable watching routine was disrupted, and learners routinely resisted transitioning into an active speaking mode.
- Cycle 2-Scaffolding Active Interaction: In Cycle 2, introducing interactive mechanisms completely altered this passive habit. By using pre-programmed pauses with embedded questions, the video structure broke up passive observation. The reflective journal highlighted that these intentional pauses built an authentic communicative need within the student cohort. The awkward silences that characterized the manual pauses in Cycle 1 were replaced by active, structured, and collaborative speaking opportunities.

4.2.2. Theme 2: Contextual relevance

The teacher’s journal highlights that personal and cultural familiarity with the visual materials served as a vital bridge for language processing, particularly for struggling learners.

- Cycle 1-Cultural Disconnect: In Cycle 1, using generic online cartoons introduced an unintended cultural barrier. For instance, during a Unit 4 lesson on food and drinks, the video featured unfamiliar Western culinary concepts such as “trifle” and “jelly”. Because these references were alien to young children living in a regional Vietnamese setting, the students struggled to draw connections between the target English vocabulary and their personal real-life experiences, causing many to disengage mentally.
- Cycle 2-The Personalization Effect: To resolve this disconnect, Cycle 2 featured customized video tours generated via AI tools to display identical, realistic representations of the students’ actual school surroundings. The journal recorded a massive shift in classroom energy when the video presented a digital model of their own school gate. Students became instantly excited, shouting, “Wow, our school gate!”. Most notably, a low-proficiency male student who had remained completely silent for the first nine weeks of the semester voluntarily raised his hand for the first time, confidently shouting, “It’s in the mountain!” when he recognized the local terrain. Familiar visuals drastically reduced extraneous cognitive load, enabling struggling students to focus limited working memory on practicing English vocabulary rather than deciphering unfamiliar cultural settings.

4.2.3. Theme 3: Psychological Safety

The final theme traces the emotional correlation between classroom anxiety and oral performance, showing how the learning medium controls the student’s affective filter.

- Cycle 1-Performance Anxiety: The journal entries from the first half of the study paint a vivid picture of communication apprehension. When asked traditional, direct comprehension questions after watching linear videos, shy or anxious students regularly displayed avoidance strategies. Several learners routinely put their heads down on their desks or actively avoided eye contact with the teacher to escape public performance errors. The fear of negative evaluation by peers acted as a severe affective filter, completely blocking oral language output.

- Cycle 2-The Safe Virtual Environment: The introduction of virtual teacher avatars completely transformed this social anxiety dynamic during the final four weeks of the project. The teacher noted that the learning environment felt exceptionally secure because students viewed the interactive digital avatar as a non-threatening, friendly partner rather than a high-stakes evaluator. Even the most anxious and quiet students were observed comfortably whispering or speaking their answers directly to the digital character on screen. Practicing with a virtual avatar provided a safe, judgment-free zone that scaffolded the students' internal confidence, empowering them to take their first essential steps toward vocalizing English phrases before interacting with real human peers or instructors.

4.2.4. Summary Matrix of Qualitative Journal Findings

To provide a consolidated reference of the thematic shifts recorded over the 12-week action research period, Table 4.3 synthesizes the core qualitative findings:

Table 4.3. Thematic Matrix of Teacher's Journal Observations

Identified Theme	Cycle 1: Linear Media Approach (Weeks 1–6)	Cycle 2: Interactive AI Approach (Weeks 9–12)	Core Pedagogical Outcome
Engagement Dynamics	Receptive viewing; high attentional focus but severe output barriers; students behaved as passive consumers.	Active generation; pause-and-predict checkpoints forced regular speaking turns and oral interaction.	Shifted students from passive observation to proactive production.
Contextual Relevance	Cultural barriers; Western concepts (e.g., <i>trifle</i>) caused processing disconnects for regional students.	Personalization breakthrough; localized visuals of their own school gate spurred immense collective interest.	Lowered extraneous load to allow low-proficiency participation.
Psychological Safety	High performance anxiety; students avoided eye contact or hid to escape public speaking errors.	Secure environment; students eagerly spoke and whispered directly to a non-judgmental AI virtual partner.	Minimised affective filter to build essential foundational confidence.

4.3. Discussion

The quantitative and qualitative results derived from this action research demonstrate that the strategic integration of interactive, AI-contextualized video lectures yielded a profound, positive transformation in the classroom learning environment. Looking at the sharp growth in the classroom observation metrics, the significant spike across the cognitive (+0.60) and behavioral (+0.60) engagement sectors directly validates the multidimensional engagement model proposed by Fredricks et al. (2004). According to their framework, genuine classroom involvement must advance beyond superficial, on-task behavioral compliance to simultaneously incorporate deep internal affect and structured cognitive investment. In Cycle 1, while traditional linear YouTube cartoons successfully captured baseline student focus, they ultimately fostered an output barrier where learners treated the educational videos as a passive TV show broadcast. This behavioral trend heavily aligns with prior empirical literature indicating that while linear media acts as an effective attention hook, it easily induces receptive watching habits rather than productive output (Cruse, 2011). Conversely, the introduction of embedded interactive checkpoints in Cycle 2 systematically disrupted this passive state, compelling young learners to utilize generative processing to actively predict, organize, and integrate upcoming linguistic concepts (Mayer, 2009). This finding surpasses generalized local claims that videos merely improve classroom energy (Hoang, 2023; Le & Tran, 2020) by empirically illustrating the distinction between the receptive enjoyment of watching a cartoon and the generative cognitive processing driven by interactive tasks.

Furthermore, the statistically significant improvement in self-reported learner attitudes ($p = 0.008$) highlights the pedagogical value of leveraging generative AI to secure localized contextualization and establish psychological safety. The exceptional quantitative performance of Item 8 ($M = 3.25$), which measured excitement regarding the AI-rendered school visuals, strongly substantiates Mayer's (2009) Personalization Principle, which states that students learn deeper when material is delivered in a highly conversational style and holds direct personal relevance. While generic Western cartoons created a cultural disconnect by presenting unfamiliar concepts like "trifle" or "jelly" to regional Vietnamese beginners, custom AI-personalization through Google Veo enabled a situated learning experience that allowed struggling students to allocate limited working memory to target vocabulary rather than decoding foreign settings. Simultaneously, the dramatic +0.45 leap in speaking confidence (Item 4) directly corroborates Krashen's (1982) Affective Filter Hypothesis. As detailed in the teacher's journal, interacting with the non-threatening AI Avatar systematically lowered foreign language performance anxiety by establishing a secure digital training zone. This pattern strongly validates Social Agency Theory, which posits that anthropomorphic social cues, such as continuous eye contact from a digital character, deepen cognitive interactions as long as the learning environment remains psychologically non-threatening (Mayer, 2014). The virtual teacher effectively acted as an essential scaffolding partner, helping anxious primary students transition safely from silent apprehension to confident oral production.

5. Conclusion

In conclusion, this action research project successfully demonstrates the significant pedagogical value of integrating interactive, AI-driven video lectures to enhance student engagement and optimize learning attitudes among primary EFL learners. The 12-week intervention revealed that standard linear video media, while highly effective at capturing baseline behavioral attention, ultimately risks inducing passive viewing habits and creating an oral output barrier when left unsupported by structured interactive tasks. By contrast, the strategic implementation of a "pause-and-predict" framework combined with custom AI-generated materials transformed the classroom dynamic, yielding statistically significant, multi-dimensional growth across behavioral, emotional, and cognitive engagement scales. Pedagogically, these findings imply that educators should transition from being passive consumers of generic, online media to active creators of customized, locally relevant digital resources. Utilizing generative AI to personalize visual assets, such as embedding familiar local school environments, drastically minimizes extraneous cognitive load and cultural barriers for young beginners, allowing them to focus restricted mental capacity entirely on target language acquisition. Furthermore, embedding structured interactive checkpoints transforms static media into a powerful catalyst for generative processing, while virtual teacher avatars provide a psychologically safe, non-threatening scaffolding space that effectively reduces foreign language anxiety and fosters communicative confidence before young learners are required to engage in real-world human interactions.

References

- Baker, C., 1992. Attitudes and language. *Multilingual Matters*.
- Braun, V., Clarke, V., 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, pp. 77–101.
- Cruse, E., 2011. Using educational video in the classroom: Theory, research and practice. Library Video Company.
- Eagly, A. H., Chaiken, S., 1993. The psychology of attitudes. Harcourt Brace Jovanovich College Publishers.
- Fredricks, J. A., Blumenfeld, P. C., Paris, A. H., 2004. School engagement: Potential of the concept, state of the evidence. *Review of Educational Research* 74, pp. 59–109.
- Guariento, W., Morley, J., 2001. Text and task authenticity in the EFL classroom. *ELT Journal* 55, pp. 347–353.

- Hoang, T. M. A., 2023. The effects of using YouTube-based videos on English speaking performance of Grade 2 students at Iris Primary School. *International Journal of Social Science and Human Research* 6, pp. 6381–6387.
- Ibrahim, M., Antonenko, P. D., Greenwood, C. M., Wheeler, D., 2012. Effects of segmenting, signaling, and weeding on learning from educational video. *Learning, Media and Technology* 3, pp. 220–235.
- Kemmis, S., McTaggart, R., 2000. Participatory action research, in “*Handbook of Qualitative Research*” (2nd ed.) N. K. Denzin, Y. S. Lincoln, Editors. Sage, pp. 567–605.
- Krashen, S. D., 1982. *Principles and practice in second language acquisition*. Pergamon.
- Le, T. T. N., Tran, D. V. D., 2020. Impacts of using cartoons as an instructional tool on English communicative skills of young learners at a Vietnamese primary school. *Ho Chi Minh City Open University Journal of Science* 10, pp. 78–86.
- Mayer, R. E., 2009. *Multimedia learning* (2nd ed.). Cambridge University Press.
- Mayer, R. E., Editor, 2014. *The Cambridge handbook of multimedia learning* (2nd ed.). Cambridge University Press.
- Minalla, A. A., 2024. Enhancing young EFL learners’ vocabulary learning through contextualizing animated videos. *Theory and Practice in Language Studies* 14, pp. 578–586.
- Nguyen, T. P. N., Pham, T. H., 2025. Video lecture design factors in STEM education. *TEM Journal*.
- Read, J. C., 2008. Validating the Fun Toolkit: An instrument for measuring children’s engagement with computing products. *Cognition, Technology & Work* 10, pp. 119–128.
- Richards, J. C., Lockhart, C., 1994. *Reflective teaching in second language classrooms*. Cambridge University Press.
- Skinner, E. A., Pitzer, J. R., 2012. Developmental dynamics of student engagement, coping, and everyday resilience, in “*Handbook of Research on Student Engagement*” S. L. Christenson, A. L. Reschly, C. Wylie, Editors. Springer, pp. 21–44.