

EFL Learners' Perceptions of AI-Generated Videos and Canva-Based Digital Worksheets: A Multi-Site Case Study in Indonesian Junior High Schools

Moh Farid Firdaus^{1*}

¹Department of English Education, Faculty of Teacher Training and Education, Universitas Islam Malang, Indonesia
(mohfaridfirdaus@gmail.com)

Dwi Fita Heriyawati²

²Department of English Education, Faculty of Teacher Training and Education, Universitas Islam Malang, Indonesia
(dwifitaheriyawati@unisma.ac.id)

Yeni Istiyani³

³Department of English Education, Faculty of Teacher Training and Education, Universitas Islam Malang, Indonesia
(istiyaniyeni1@gmail.com)

*corresponding author

Abstract: *The rapid adoption of AI-generated videos and Canva-based worksheets in Indonesian EFL classrooms has not been matched by an empirical understanding of student experiences. This study highlights three gaps in the literature: these two media are generally studied separately rather than together; affective and metacognitive dimensions remain less explored than cognitive outcomes; and few studies compare students' perceptions of these media across public, private, and Islamic boarding schools. Using a multi-site interpretative qualitative case study design, this research examines the perceptions of seventh-grade students in these three types of schools, with nine students purposively selected, three from each school. Data were collected through semi-structured interviews, classroom observations, and documentation, then analyzed using the interactive model of Miles, Huberman, and Saldaña (2014) with NVivo 12. The findings show five patterns: consistent perceptions of usefulness across locations; perceptions of ease of use covering both intrinsic (interface) and extrinsic (infrastructure) dimensions; learning motivation developing through different pathways at each location; student engagement reflecting each institution's typology; and learner autonomy shaped by institutional epistemology. This study proposes a Contextually Mediated Technology Acceptance and Engagement Model (CM-TAEM), an integrative framework that places institutional culture and digital infrastructure equity as moderators of technology acceptance and engagement. The results suggest that integrating AI-based media in EFL classrooms should take into account the specific conditions of each institution, rather than assuming uniform effectiveness.*

Keywords: *AI-generated videos; Canva-based digital worksheets; EFL learners; Junior high schools; students' perceptions*

Article History:

Received : 03 July 2026

Accepted

: 26 September 2026

Revised : 21 September 2026

Available Online

: 30 September 2026



Copyright © 2026 The Authors. The journal uses an Open Access policy under a Creative Commons Attribution-NonCommercial 4.0 International License

Suggested Citation Format:

Firdaus, M. F., Heriyawati, D. F., & Istiyani, Y. (2026). EFL Learners' Perceptions of AI-Generated Videos and Canva-Based Digital Worksheets: A Multi-Site Case Study in Indonesian Junior High Schools. *PANYONARA: Journal of English Education*, 8(2), 355–380. <https://doi.org/10.19105/panyonara.v8i2.25528>.

INTRODUCTION

Digital technology has become a standard component in foreign language classrooms worldwide, and its growth in Indonesian secondary education has been further accelerated by post-pandemic digitalization and the Merdeka Belajar curriculum reform (Kemendikbudristek, 2022; UNESCO, 2023). In this context, two media are quite popular in EFL classes in Indonesia: Artificial Intelligence (henceforth AI)-created learning videos and Canva-based digital student worksheets (LKPD). Research on each of these media has advanced rapidly, but to understand what is known and what remains unclear, it is necessary to integrate both bodies of research rather than treat them as a single literature.

Studies on AI-generated videos in EFL settings show a fairly consistent picture: video content that combines verbal and visual channels supports understanding and memory more effectively than input from a single channel Reynolds et al. (2022), students' acceptance is influenced by perceptions of usefulness and ease of use Zhang et al. (2024), AI-created videos can approach the effectiveness of human-made videos in teaching certain grammar constructions Kim (2024), and conversational AI videos boost situational engagement in language centers (Phi & Khanh, 2024). So, this branch primarily positions AI videos as a medium for understanding and acceptance.

Parallel literature considers Canva-based digital worksheets as a different object of study, emphasizing design quality and visual engagement: students find materials designed with Canva more visually appealing compared to conventional paper worksheets (Savitri & Ratri, 2025), Canva supports interactive and multimodal e-modules (Fitria, 2024; Purwasi et al., 2023), and independent use of interactive worksheets is linked to more self-directed learning behavior (Ly, 2024; Mutiaraningrum & Nugroho, 2021). More broadly, EFL technology research connects AI-supported teaching and structured autonomy support with motivation, engagement, and more consistent self-directed or independent learning (Wei, 2023; Chiu et al., 2023; Fujii, 2024; Al-Khasawneh et al., 2024).

Looking at this literature, there are still four shortcomings. First, there is a fragmentation issue: AI videos and Canva worksheets are almost always studied separately Phi & Khanh (2024) and Kim (2024) only looked at AI videos, Savitri & Ratri, (2025), Fitria (2024) and Purwasi et al. (2023), only studied Canva worksheets, and even studies that touch on both (Ly, 2024) do not treat them as interconnected stages in a single lesson. Looking through Mayer's Cognitive Theory of Multimedia Learning (2021), this separation turns into a substantive limitation, not just a convenience: AI videos relate to the comprehension phase in multimedia processing and Canva worksheets to the application

and consolidation phase, so the cycle of acceptance and motivation connecting the two TAM Davis (1989) and SDT Deci and Ryan (2000) cannot be observed if only one medium is studied, and how acceptance at the comprehension stage continues to worksheet-based engagement is still largely unexplored. Second, there is a results gap: in the video-AI literature, acceptance and understanding are discussed more often, whereas design quality and visual engagement are more prominent in the Canva literature. Though motivation, engagement, and autonomy have been explored more widely in the EFL technology literature, they are mostly phenomena that occur outside the specific video-AI and Canva worksheet literature. Thus, little research has examined the long-term effects of combining the two media on motivation, engagement, and autonomy.

Third, there is an institutional gap. The existing research in Indonesia on digital media has mostly been conducted in vocational higher education (Mutiaraningrum & Nugroho, 2021). It does not distinguish between types of schools at the middle school level, thus it is still unclear how AI videos and Canva worksheets are received in public, private, and pesantren-based schools, which make up a large part of Indonesia's diverse secondary education landscape. Pesantren-based schools, which combine religious knowledge and unique community structures with the same national curriculum, receive very limited attention. Fourth, in line with institutional gaps, there is a theoretical-integration gap. Institutional culture has long been said to influence how members of an organization adopt new practices Bourdieu (1977) and Hofstede (2001), referred to here as organizational culture theory background rather than direct empirical support, and cross-country research confirms that technology use in the classroom is influenced by institutional and cultural contexts Grothaus (2023); however, the relationship between institutional culture, digital infrastructure equity, and technology adoption still receives limited empirical attention in Indonesian EFL classrooms, and meta-analytic evidence based on TAM about AI in education Xue et al. (2026), synthesizes usefulness and ease-of-use predictors without considering the institutional conditions in which these perceptions are formed. Because learner autonomy has mostly been studied as an individual-level concept (Fujii, 2024; Al-Khasawneh et al., 2024), rather than as something that might vary depending on institutional epistemology, the relationship between technology acceptance, engagement, and learners' autonomy in different institutional contexts is still fragmented and has not been analyzed as a single complete system.

This study directly and sequentially addresses all four of these gaps. It looks at AI-created videos and Canva-based worksheets as a single, sequential learning package, rather than separate interventions. To tackle the fragmentation gap, it analyzes perceptions across five interconnected dimensions, usefulness, ease of use, motivation, engagement, and learner autonomy, rather than just focusing on cognitive outcomes or acceptance. To address the outcome gap, it compares three types of deliberately selected schools (public, private, and pesantren-based) using a multi-site case study design, directly addressing the

limited attention to the pesantren context. To close the gap in theoretical integration, this study develops a Contextually Mediated Technology Acceptance and Engagement Model (CM-TAEM), which introduces institutional culture and digital infrastructure equality as contextual moderators along with constructs like Extrinsically Perceived Ease of Use, Relational Channel Specification, Institutional Epistemology, and Institutional Engagement Orientation operationalizing. In the EFL context, the theoretical relationship between institutional context and technology acceptance is suggested by organizational culture literature, but not empirically tested (Grothaus, 2023).

Based on this gap, this study asks: (1) How do seventh-grade students perceive the benefits and ease of use of integrating AI videos and Canva worksheets in EFL learning? Moreover, (2) How does this affect motivation, engagement, understanding, and learning independence, as well as how does the context of different institutions moderate these perception patterns? The study contributes in three ways. Contextually, this investigates an integrated package of AI videos and Canva worksheets across three types of institutions, including pesantren. Methodologically, this uses a multi-site case study design that allows for comparisons across locations rather than descriptions of just one Site. Theoretically, this extends TAM, CTML, SDT, and the Learner Autonomy Framework into an integrated CM-TAEM model.

LITERATURE REVIEW

This study draws on four theoretical perspectives that together form a conceptual framework for understanding students' perceptions of AI-generated videos and Canva-based digital worksheets across four dimensions: motivation, engagement, understanding, and learning independence. Each perspective is also explained explicitly below in relation to the previously identified research gaps.

Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM; Davis, 1989), later expanded by Venkatesh et al. (2003) and specifically applied to AI-based educational tools by Zhang et al. (2024) and at the meta-analysis level by Xue et al. (2026), argues that perceived usefulness and perceived ease of use are the main determinants of technology acceptance. This study uses TAM as a lens to understand students' evaluative and attitudinal responses to AI videos and Canva worksheets, while expanding the model through the Extrinsically Perceived Ease of Use construct proposed in CM-TAEM, to consider ease-of-use perceptions shaped by infrastructure conditions beyond students' control, a dimension that, in meta-analytic TAM literature, Xue et al. (2026) do not distinguish by institutional context.

Cognitive Theory of Multimedia Learning (CTML)

Mayer's (2021) Cognitive Multimedia Learning Theory explains how presenting verbal and visual information simultaneously supports the formation of mental models, and it has been extended to AI-generated content by identifying new boundary conditions for AI-generated multimedia (Lim et al., 2023; Kim, 2024). This study uses CTML to place AI-generated videos in the understanding phase of the learning sequence examined here, functionally distinguishing them from the role of application and consolidation played by Canva worksheets—a difference that underpins the fragmentation gap discussed earlier.

Self-Determination Theory (SDT)

Self-Determination Theory (SDT), drawing on Deci and Ryan (2000), explains motivation through the fulfillment of psychological needs for autonomy, competence, and relatedness, and have been used to explain how AI-based instruction maintains L2 motivation and self-regulated learning Wei (2023) as well as how basic psychological needs operate under various learning conditions (Vansteenkiste et al., 2020). This study uses SDT to interpret motivational pathways through feedback-based interactivity on Canva worksheets that might fulfill competence needs developed in the previous video-based understanding stage, a sequential mechanism identified as underexplored in the outcome-dimension gaps mentioned above.

Learner Autonomy Framework

Benson's (2011) Learner Autonomy Framework, along with related research on the use of technology that supports autonomy (Wei, 2022; Fujii, 2024; Al-Khasawneh et al., 2024), sees autonomy as a student's ability to take control of their own learning process. Ly (2024) reported that EFL students who used digital worksheets exhibited more independent learning behaviors and a greater sense of control over their learning process. However, this relationship has not been studied across different types of schools where institutional epistemologies might affect how much autonomy is encouraged or allowed for students. This study uses the Learner Autonomy Framework, along with the constructs of Institutional Epistemology and Institutional Engagement Orientation proposed in CM-TAEM, to directly examine the effects of these conditions.

Together, these four perspectives form an integrated conceptual framework: TAM explains how students accept and assess technology (perception dimension); CTML explains how multimedia design affects cognitive processing (understanding dimension); SDT explains the sources and quality of learning motivation (motivation dimension); and the Learner Autonomy Framework explains how technology can empower or limit student learning autonomy (independence dimension). Engagement is treated as a behavioral manifestation of intrinsic motivation that develops within the context of effective multimedia design (Fredricks et al., 2004). It is thus understood as the result of the

interaction among TAM, SDT, and CTML, rather than as a fifth standalone construct. The CM-TAEM model proposed in this study expands this integration by adding institutional culture and digital infrastructure equality as moderating conditions- a layer of context that, according to the gap identified above, existing literature has not included yet.

METHOD

Research Design

This study used an interpretative qualitative approach with a multi-site case study design. The interpretative paradigm was chosen because the research focuses on students' subjective meanings when interacting with digital learning media, a complex phenomenon that cannot be reduced to measurable variables (Creswell & Poth, 2018; Merriam & Tisdell, 2016). The multi-site case study design is based on the logic of replication (Yin, 2018); each school is treated as a standalone case (within-site analysis) and then systematically compared (cross-site analysis) to identify similarities and differences across contexts. This design allows for analytic generalization, developing theoretical understanding that can be transferred to similar contexts rather than statistical generalization (Stake, 1995; Yin, 2018).

Research Location and Participants

The research was conducted at three middle schools selected purposively based on the principle of maximum variation sampling (Patton, 2015), with different institutional characteristics as summarized in Table 1.

Table 1.
Profile of Participating Schools

Pseudonym	Types	Digital Infrastructure	Main Characteristics
SMP Negeri Harapan Bangsa (Site A)	Public school	Moderate, varied	Students come from various socio-economic backgrounds; technology is adjusted based on availability; there are approximately 210 students; classes consist of 30 students each.
Darussalam Integrated JHS (Site B)	Islamic boarding school	Limited, controlled	The school follows a national and religious curriculum; access to technology is strictly regulated; there are approximately 180 students enrolled; and each class has about 30 students.
Mandiri Prestasi JHS (Site C)	Private school	Adequate, well-established	Established digital infrastructure; high level of digital literacy; approximately 100 students; 18 students per class; predominantly upper-middle-class.

The participants consisted of nine seventh-grade students (three per school), selected based on three criteria: (1) they had attended both the AI video learning sessions and the Canva worksheet sessions, each held twice at each school; (2) they actively completed the Canva worksheets immediately after each of the two AI video sessions, instead of just passively watching the videos; and (3) they had at least 80% attendance during the studied semester. This balanced frequency aligns with the CTML-based reasoning explained in the Introduction: each AI video session, which supports the understanding phase, was immediately followed by a Canva worksheet session, which supports the application and consolidation phase, so both media formed a paired understanding-to-production sequence at each school. All participants were coded P1-P9 (P1-P3: Location A; P4-P6: Location B; P7-P9: Location C).

These criteria were adopted so that every participant receives equal and complete exposure to both media, which is important given that the goal of this study is to examine how the two media are perceived as a unified sequence rather than as separate experiences. However, as is well known, the prerequisite of completion (active completion), along with the attendance requirement (at least 80%), will likely rule out less engaged students, those who attend infrequently, those who struggle academically, and/or those with limited access to interventions due to infrastructure constraints. Therefore, the sample reflects more the attitudes of the students who are consistent and capable of engaging with both media, and the findings should not be interpreted as the entire spectrum of attitudes within each class, including those who may not have participated in the activities or may have experienced difficulty with the intervention. These limitations are also discussed further in the Conclusion. The total of nine participants was determined based on two methodological principles: information power (Malterud et al., 2016) and replication logic (Yin, 2018), which are elaborated together with the study's saturation assessment, since saturation was evaluated once data collection and initial coding were underway, rather than fixed in advance.

Intervention Media Description

Two types of digital media are used in an integrated way. The AI-generated instructional videos are created with the help of the ElevenLabs text-to-speech system, which provides an English voice for the video, CapCut for video editing, and illustrative videos and on-screen text. The videos, which are 5-8 minutes long, are divided into three sections: an introduction to the topic's context, a teaching and learning segment, and a closing segment. Digital student worksheets (LKPD) are created using Canva for Education and consist of 8x3 activities that are progressively planned based on the scaffolding principle from basic comprehension activities (multiple-choice based on text and image) to production with open-ended questions. Full details of both media are presented in Table 2.

Table 2

Description of Learning Media Utilized

Aspect	AI-generated instructional videos	Canva-Based Digital Student Worksheet
Material	About Me (English for 7th Grade Junior High School)	About Me (English for 7th Grade Junior High School)
Platform	ElevenLabs (AI narration) + CapCut (editing)	Canva for Education
Function	To present learning material in an audiovisual format	To facilitate student practice, reflection, and comprehension reinforcement
Main Components	AI-based English narration, animation, visual illustrations, supporting text, and transitions	Activity instructions, vocabulary exercises, comprehension questions, sentence construction tasks, and reflective activities
Design Principles	CTML: signaling, segmenting, personalization (Mayer, 2021)	Student-centered learning; gradual scaffolding
Duration	5-8 minutes per learning session	15-20 minutes to complete each worksheet
Access Method	Played via LCD/projector in class; Google Drive link for individual access (Sites A & C)	Accessed via link or QR code using a smartphone or computer
Frequency	2 learning sessions	2 learning sessions (consolidation point immediately after each video session)

Table 2 summarizes two media as one sequential learning package, not two separate tools: AI-generated videos aligned with the understanding phase in CTML (Mayer, 2021) through a mix of narration, animation, and visual illustrations, while Canva-based worksheets align with the application and consolidation phase through progressively challenging tasks completed right after each video. When read together, the platform, function, and frequency rows show that each AI video session is paired directly with a Canva worksheet session, so understanding and production are experienced as a single continuous cycle rather than separate, unrelated activities.

Instrument Validation

Semi-structured interview guides and observation protocols were developed based on a literature review and conceptual framework. Content validity was evaluated through expert review by two experienced researchers in English language education and educational technology—one expert in EFL learning design and the other in digital learning media—who assessed each item according to four theoretical constructs of the study (TAM, CTML, SDT, and Learner Autonomy) for relevance and clarity; their feedback led to revisions in word choice and construct order in the questions. Separately from this validity assessment, the instruments' readability was tested through pilot interviews with two

students outside the sample, who evaluated the clarity and ease of understanding of phrasing, not construct validity, resulting in clarification of terminology in three questions.

Data Collection

Data was collected using three integrated techniques to achieve methodological triangulation (Yin, 2018). First, individual semi-structured interviews (35-50 minutes per session, conducted in Indonesian) were transcribed verbatim; to reduce social desirability bias, all interviews were conducted by an independent interviewer with no direct teaching relationship to the participants. The accuracy of the transcripts was confirmed through member checking; three of nine participants provided minor clarifications, which were incorporated into the final transcripts. Second, structured classroom observations were conducted in two sessions per school (a total of six sessions across three locations), each lasting 80 minutes, representing a different data collection procedure and not meant to match one-to-one with the media exposure schedule reported above (two AI video sessions and two Canva worksheet sessions per participant, following the order AI Video Session 1 → Canva Worksheet Session 1 → AI Video Session 2 → Canva Worksheet Session 2).

From the two observation sessions per school, one was scheduled during the AI video session and the other during the paired Canva worksheet session, so both types of media could each be observed at least once at every site, rather than observing just one type of media; these six observation sessions therefore serve as a sample, not the full coverage, of the twelve media exposure sessions (four per site) experienced by each participant. Each observation session followed a protocol that recorded engagement patterns, student responses, class interactions, and technical issues, and this observation data was mainly used to adjust and provide context for interview results—for example, to support self-reported claims of motivation or ease of use with observable behavioral indicators (Results, RQ1-RQ2), rather than serving as the only source of data. Field notes were verified by peer observers during both sessions at each school. Third, the documentation includes lesson plans (RPP), Canva student worksheets (LKPD), AI video scripts, and school technology policy documents.

Data Analysis

Data were analyzed using Miles, Huberman, and Saldaña's (2014) interactive model in NVivo 12, following Saldaña's (2013) First- and Second-Cycle coding procedures rather than grounded-theory axial coding. There were 187 initial codes, grouped into 24 categories through second-cycle open, descriptive coding, and then further grouped into five main themes through thematic grouping of the categories. In addition to the trustworthiness strategies listed in Table 3, a consistency check (Cohen's $\kappa = 0.82$; substantial agreement; Landis & Koch, 1977) was used to check reliability on 20% of the transcripts, but as a

consistency check to the study's interpretative paradigm rather than as a positivist validity criterion.

Analysis was done in two stages, a case-to-case study (Yin, 2018). Independent thematic descriptions were first generated within each school site for each of the five constructs (perceived usefulness, perceived ease of use, motivation, engagement, and autonomy). The descriptions were then cross-analyzed to determine convergence (patterns found in all three sites) and divergence (site-specific patterns), and interpreted in relation to the school's infrastructure and institutional culture (Table 1), rather than as context-free patterns. This case analysis → cross-case comparison → pattern identification → contextual interpretation sequence produced the cross-site and case-specific patterns reported in the Results and synthesized in Table 4.

Data Saturation and Transferability

Data saturation was assessed using the principle of information power (Malterud et al., 2016) rather than a fixed sample-size rule, combined with the replication logic underlying the multi-site design (Yin, 2018). Interviews were coded in order (P1-P9: Site A, P1-P3; Site B, P4-P6; Site C, P7-P9). The codebook had stabilized, with no new codes identified during coding of the seventh interview (P7), but P8 and P9 were also fully coded and compared against the codebook, rather than assuming that no new codes would be identified later. The three participants per site were not an independent statistical sample, but rather a somewhat "rich" sample with a variety of structures, adequate for within-site analysis and cross-site comparison.

The study, therefore, does not seek statistical generalization but analytical transferability, reader judgments of the findings' relevance to their context based on a detailed description. The three types of schools reflect the typical types of junior high schools in Indonesia: public schools that prioritize academic achievement, pesantren schools that emphasize the role of teachers, and private schools that have abundant digital facilities. To substantiate more general claims, larger multi-site or multi-method studies are required.

Reflexivity and Trustworthiness

As a practitioner-researcher who also designs the media under study, the potential for confirmation bias and power differentials is managed through: (1) using an independent interviewer; (2) maintaining a reflective journal after each session; (3) conducting two peer debriefing sessions; (4) performing negative case analysis; and (5) maintaining a comprehensive audit trail (Creswell & Poth, 2018). The trustworthiness of the research is ensured through Lincoln and Guba's (1985) four criteria, as summarized in Table 3. It summarizes how each of the four reliability criteria from Lincoln and Guba (1985) was

operationalized in this study, linking each criterion to the specific strategies used and their application during data collection and analysis.

Table 3.
Research Trustworthiness Strategies

Criterion	Strategy	Implementation
Credibility	Methodological triangulation, member checking, peer debriefing, and negative case analysis.	Three data collection techniques: transcript confirmation by 9 participants, two peer debriefing sessions, and explicit documentation of negative cases.
Transferability	Thick description	In-depth description of school profiles, participant characteristics, and institutional context.
Dependability	Audit trail, inter-rater reliability	Documentation of methodological decisions; independent coding used as a supplementary consistency check (Cohen's $\kappa = .82$), interpreted alongside, not in place of, triangulation, member checking, and peer debriefing.
Confirmability	Reflexivity, audit trail	Reflective journal; documentation of the analysis process; researcher positionality statement.

As shown in the table, credibility and confirmability relied more on triangulation and reflective documentation. At the same time, dependability treated inter-rater coding more as an extra consistency check than as an independent validity claim.

RESULTS AND DISCUSSION

This section presents the findings from semi-structured interviews, classroom observations, and documentation involving nine seventh-grade students across three sites: a public junior high school (Site A), a pesantren-based junior high school (Site B), and a private junior high school (Site C). Participants are coded P1-P3 (Site A), P4-P6 (Site B), and P7-P9 (Site C). The study's two research questions organize findings; their theoretical interpretation follows in the Discussion.

RQ1: Students' Perceptions of the Usefulness and Ease of Use of AI Videos and Canva Worksheets

Perceived usefulness. All nine participants across the three sites viewed the AI-generated videos and Canva worksheets as useful. At Site A, P1 explained that the videos make lessons easier *"because I can see and hear the explanation at the same time, which is clearer than just reading the book,"* and P3 said the videos let students *"have more knowledge and gain more than what the book offers,"* both framing the usefulness comparatively against

the textbook. At Site B, P5 said it was *"really good and enjoyable to watch"* (an affirmative but non-comparative answer). At Site C, P9 correlated usefulness with curricular fit: *"The material is what is being taught in class, so it is not confusing,"* and P8 highlighted practical support with vocabulary. This consistency in participants' responses suggests that perceived usefulness was consistently high across sites, but with different anchors, specifically: comparative academic gain at Site A, affective comfort at Site B, and curricular fit at Site C.

Perceived ease of use. Unlike usefulness, ease of use diverged sharply by Site. At Site A, P1 described the media as *"quite easy, just type, just click around,"* and P3 reported no obstacles to usage. At Site B, P5 attributed difficulty to connectivity: *"the signal here is unstable, so sometimes the pages take forever to load, which can be kind of distracting,"* even though P4 and P6 both called Canva's interface itself easy and engaging. At Site C, P7 found the worksheet manageable without instructions, while P9 added a qualifier, *"easy to use, but if the internet connection is slow, you have to wait a bit."* Classroom observation confirmed students accessed the media via shared school tablets. Read together, this evidence indicates that perceived ease of use has two distinguishable sub-dimensions: intrinsic PEU (ease inherent in interface design, high at all three sites) and extrinsic PEU (ease conditioned by infrastructure beyond students' control, ranging from largely absent at Site A, to moderate at Site C, to severe at Site B). The contrast is sharpest between Site A and Site B: both report high intrinsic PEU, but only Site B's extrinsic PEU is constrained by signal quality and device policy.

This means that a negative case at Site A qualifies this picture. Although P1 did not report any technical difficulties, he shared his opinion that the worksheets lacked stimulating elements, noting that they *"would be even better if a little fun game or icebreaker was added."* This suggests that ease of use and quality of engagement are statistically separable factors. Although PEU may not be an issue, engagement may be for students with a more active, gamified orientation.

RQ2: Motivation, Engagement, Understanding, and Learning Independence, and Institutional Moderation

Learning motivation. All nine participants reported increased motivation linked to the multimodal format, but its psychological source diverged by Site. At Site A, P2 linked motivation to competition: *"I'm more motivated because I can see my score immediately and compare it with my friends, so I want to keep learning to do better."* At Site B, P5 tied motivation to comfort (*"it helps me feel motivated and calm to study"*), while P6 framed it epistemically (*"because I want to know a lot more"*). At Site C, P9 linked motivation to peer collaboration (*"it's more fun because we can work on worksheets with friends, so it doesn't feel stressful"*), while P8 linked it to reduced pressure relative to printed exercises. Several participants spontaneously contrasted the media with textbook-only instruction in

motivationally inferior terms; P1 called such lessons *"less interesting, and there is also a chance of misunderstanding the material."*

A theoretically significant negative case emerged at Site B: P4, P5, and P6 independently stated that the teacher's role could not be replaced by digital media, P4 explaining simply, *"No, because teachers are more important."* *Interpretation.* This cohesive response cluster, distinct from Sites A and C, suggests a pesantren pedagogical orientation in which the teacher-student relationship extends beyond knowledge transfer. The observation results confirmed the motivation pattern revealed by the self-reported data: at all three sites, the students were highly enthusiastic and reluctant to finish the activities, while at Site C the observation data showed that students talked to one another more often when working on the worksheets, and that more often they completed the tasks according to the teacher's reflection journal, which is an independent behavioral indicator that supports the motivational findings.

Student engagement. The main driver of engagement, albeit with varying characteristics, was the real-time feedback from the Canva worksheets. At Site A, P3 spoke about competitive-evaluative engagement: *"When the score is instant, there's a need to correct what I'm doing; it gets me thinking more."* P7 at Site C referred to reflective-metacognitive engagement, replaying the unclear parts of the videos *"to know which vocabulary I still need to study before the test,"* and P9 at Site C used the results from the worksheets to decide what to revisit. The majority of the participants commented on increased focus connected to the audiovisual nature of the media, but Site C's was more dependent on length: P7 talked about how *"as long as the video isn't too long, I get tired too,"* and P9 indicated that modality variation was a factor to increase focus.

As for exposure intensity, two complementary negative cases were identified: the participant of P7 (Site C) found that focus was reduced when the exposure intensity was high, and the participant of P5 (Site B) found that the exposure intensity was low and that focus was reduced. This suggests, but does not prove, that intensity of exposure and engagement may not be linear, and is a hypothesis-generating observation, based on two contrasting cases in a nine-participant design, not a clear inverted-U curve, systematic tracing of which is left to future research. In addition to the differences between the three sites, engagement orientation was fairly consistent across sites (competitive-evaluative at Site A; affective-relational at Site B; reflective-metacognitive at Site C), suggesting that engagement form is not solely determined by media quality.

Material understanding. Multimodal aspects (image, animation, audio) were reported to support comprehension at all locations, most explicitly by P9, who reported that the grammar animation helped him: *"the animation displays the sentences one by one and I understand the pattern better than from the formulas alone."* At Site A, P2 explained that the combined images, audio, and animation made the content *"more complete"* and *"more detailed"* than any single channel. At Site C, the participants were more specific: P7

emphasized the need for narration in addition to pictures *"it's confusing some parts if it's just pictures and not someone who can explain it,"* P8 noted the voice and text combination, and P9 cited animation as a tool to help illustrate grammar patterns.

A negative case that has repeatedly occurred is audio quality at Sites A and C: P1: *"Sometimes the background music is louder than the narrator's voice, so I have to concentrate extra, and there are still parts I can't hear"* and P3: *"There is an articulation problem too, I have to read on the screen when I cannot understand what the person is saying.* Mild but similar problems were observed at Site C. Audio quality was not mentioned by any Site B participant. The lack of audio complaints at Site B might be due to an evaluative emphasis on aesthetic appreciation, but not necessarily a lack of the problem at Site B. As with Sites A and C, the pattern of observation data confirmed that when the audio was not clear, students at Site A were observed as frowning and looking to their deskmates, while students at Site C were actively re-playing segments, a remediation strategy not observed at any of the other sites, which is indicative of a more advanced metacognitive disposition in the students at Site C rather than a difference in the underlying (identical) material.

Learning independence (autonomy). Autonomy showed the sharpest divergence among the four constructs, distributed across three qualitatively distinct patterns. At Site A, P2 explained, *"I want to be able to use English for my future job, so learning with videos at home helps me keep studying even when there are no classes"*; P1 similarly connected independent learning to a career aspiration. At Site C, orientation toward independence was conditional and semi-collaborative: P7 explained, *"I can study on my own with videos, but I prefer having a friend to discuss with when there's something I don't understand"*; P8 described a similar pattern, turning to the teacher when material proved difficult; P9 pointed to an infrastructural barrier distinct from Site B's ideological one: *"the Canva worksheets can be done on your own outside class, but the videos need a good internet connection, which we don't always have at home."*

At Site B, all three participants stated they did not study independently using the media outside class. P4 cited dormitory restrictions; P5 added a self-regulation concern (*"when I study alone, I lose focus. If it's with the teacher, I understand faster"*); and P6 emphasized the teacher's irreplaceability: *"at our boarding school, we're not allowed to use phones freely, and even if we could, I think learning with a teacher is still more effective than learning on your own from videos."*

At Site A, both participants constructed independence as freedom from reliance on any single instructional source—the most optimistic orientation toward independent media use among the three sites. Site C's orientation was more open than Site B's but more cautious than Site A's. At Site B, participants' responses go beyond the structural constraints of device restrictions to articulate an epistemological belief in the teacher's superiority; this cohesive negative-case cluster indicates that the absence of self-directed learning there reflects structural, ideological, and technical barriers operating together, rather than a lack

of interest in the media itself. The barriers at Site B and Site C are not a continuum, but are categorical: whereas Site B's are structural-ideological and are unlikely to be overcome by technology access alone, Site C's are technical-situational and may be addressed through infrastructure improvements; thus, Site C occupies a transitional space between the reliance of Site B and the aspirational autonomy of Site A.

Cross-site synthesis. When read together, RQ1 and RQ2 reveal an overall pattern of strong convergence at the level of underlying perceived pedagogical value (nearly universal acceptance, motivational and comprehension benefits), and systematic divergence in the underlying mechanisms and manifestations, consistently patterned by institutional context: competitive-individual at Site A, relational-affective at Site B, and reflective-collaborative at Site C (see Table 4). This alignment-divergence pattern recurs across all four theoretical frameworks (TAM, SDT, CTML, and Benson's autonomy framework), providing convergent support for the role of institutional context; its theoretical implications are developed in the Discussion.

Table 4

Synthesis Findings Matrix Across Five Themes and Three Sites

Theme	Site A (Public Junior High)	Site B (Islamic Boarding School)	Site C (Private Junior High)
Technology Acceptance (TAM) PU & PEU	High & elaborative-comparative PU; high intrinsic PEU; no extrinsic barriers	High & affirmative-concise PU; extrinsic PEU severely constrained (signal & device policy)	High & pragmatic-curricular PU; moderate PEU, conditioned by institutional infrastructure
Learning Motivation (SDT)	Competitive-comparative; driven by score visibility & peer comparison; autonomy & competence dominant	Emotional-relational; teacher irreplaceable; relational channel specificity active	Instrumental-social-outcome-oriented; motivation based on relevance & peer collaboration
Student Engagement (SDT + CTML)	Competitive-achievement-oriented; real-time feedback triggers intensive cognitive engagement.	Affective-relational; decline in focus with limited exposure (inverted U-curve from below)	Reflective-metacognitive; sensitive to duration & modality variation; calibrated engagement
Content Comprehension (CTML)	Synergistic multimodal (image+audio+animation); technical audio barriers critiqued in detail & constructively	Strong appreciation of audio-visual aesthetics; audio barriers not articulated technically	Selective-contextual; subtitles & vocabulary annotations proposed; active and independent replay strategy
Learning Autonomy (Benson, 2011)	Aspirational autonomy; explicitly linked to long-term career goals	Structural heteronomy: ideological, structural, and technical barriers	Conditional-semi-collaborative autonomy;

		operating simultaneously	transitional position; technical-situational barriers
Emergent Contextual Dimensions	Institutional engagement orientation: competitive- achievement	Relational channel specificity; institutional epistemology; institutional engagement orientation: affective- relational	Institutional engagement orientation: reflective- metacognitive; scaffolded autonomy (transitional)
CM-TAEM Moderators	Layer 1: academic achievement & aspirations. Layer 2: adequate infrastructure	Layer 1: teacher- centric & transmission epistemology. Layer 2: limited infrastructure	Layer 1: pragmatic & collaborative. Layer 2: moderate infrastructure, dependent on school facilities

Table 4 brings together the findings from each Site across all five themes: the Technology Acceptance and Learning Motivation rows show that all three sites converged on positive outcomes but diverged in the psychological mechanisms or infrastructure conditions involved. In contrast, the final two rows summarize site-specific constructs and the CM-TAEM moderator layers proposed to explain those differences.

DISCUSSION

Interpreting Perceived Usefulness and Ease of Use through TAM

This pattern of findings invites a critical reading of the Technology Acceptance Model (Davis, 1989), which positions PU and PEU as the two main determinants of behavioral intention to use technology. The high consistency of PU across sites is compatible with TAM's central premise that the instrumental relevance of technology strongly predicts acceptance regardless of institutional context, and is further supported by recent meta-analytic evidence that PU consistently exerts a stronger influence than PEU on technology acceptance in AI-supported educational contexts (Xue et al., 2026).

The PEU findings, however, challenge TAM's original treatment of the construct as purely an individual-cognitive evaluation of interface complexity. In an educational context with substantial infrastructure disparity, as is common in Indonesia, PEU needs to be decomposed into intrinsic and extrinsic dimensions. The extrinsic PEU construct proposed here is conceptually close to perceived behavioral control in the Theory of Planned Behavior (Ajzen, 1991) and to the facilitating-conditions dimension in UTAUT (Venkatesh et al., 2003), but differs in that it treats infrastructure as a moderator that operates independently of individual perception.

These findings are consistent with Savitri & Ratri (2025) regarding the dominance of Canva's visual interface in facilitating acceptance of digital LKPD, and with Zhang et al. (2024) and Fitria (2024) regarding the contribution of AI-generated videos to instructional

clarity. Unlike those studies, which were generally conducted within a single, relatively homogeneous type of school, this multi-site design allows PEU variation to emerge that a single-site design would obscure. Wei (2023) similarly highlights infrastructure conditions as a factor moderating the effectiveness of learning technology in developing-country contexts, supporting the case for extending TAM with an infrastructure-equity dimension as a contextual construct.

The negative case of Site A's criticism of insufficient gamification indicates that successful instructional adoption requires more than meeting a PEU threshold, and points to learning-experience quality as an additional moderator relevant to TAM in the junior-high EFL context. Practically, developers of digital LKPD should consider (1) a low-bandwidth or offline mode for pesantren settings, addressing extrinsic PEU; (2) moderate gamification elements to meet stimulation needs, particularly at Site A; and (3) production standards that support use without requiring prior procedural instruction, across contexts.

Interpreting Motivation, Engagement, Understanding, and Autonomy, and Their Institutional Moderation

Motivation (SDT). Motivation patterns can be integrated into a coherent framework provided by SDT (Deci & Ryan, 2000). According to SDT, the internal motivation is a function of three basic psychological needs: Autonomy, Competence, and Relatedness. There seems to be a need for competence, as indicated by the instant visual feedback afforded by the Canva worksheet, which is consistent with the accounts of the Site A participants' sense of satisfaction and visibility of achievement; there seems to be a relative need for autonomy, at least at Site A, and more contingently at Site C.

The most theoretically valuable nuance is given by relatedness. Relational needs were addressed at Site C through technology-supported collaboration among learners. This negative case at Site B suggests that the needs for relatedness could not be transferred from the human channel (the teacher) to the technological channel, but rather stayed with the teacher, thus illustrating relational channel specificity: a tendency to satisfy connectedness needs via culturally specific institutional channels, not interchangeable with technological channels. This construct has an affinity with the relational dimension of learning culture in Hofstede's (2001) framework, but is operationalized here specifically in the context of learning-technology acceptance.

These findings extend Zhang et al. (2024) and Kim (2024), who confirm the positive contribution of AI audiovisual elements to motivational engagement but do not test their interaction with institutional relational structures. Vansteenkiste et al. (2020) argue that differences in learning context shape motivation by supporting or constraining the fulfillment of basic psychological needs; consistent with this, Site C students appeared motivated by reduced learning pressure and greater perceived control, whereas Site A students were more strongly motivated by competitive academic achievement. Within

control-value theory, Pekrun et al. (2026) similarly emphasize that learning emotions are mediated by perceptions of control and value, both of which are strongly shaped by institutional context, helping to explain why Site C reported control-based motivation while Site A reported value- and achievement-based motivation. In practice, the results indicate that AI videos could be used as initial material for teachers to follow up on later, structured competition in worksheets could be optimized in public junior high schools, and small-group tasks could be designed to increase social-collaborative motivation in private junior high schools.

Engagement (SDT + CTML). Cognitive engagement driven by immediate feedback can be understood in terms of skill needs according to SDT (Deci & Ryan, 2000): instant, informative feedback provides a clear sense of mastery that sustains cognitive engagement. Meanwhile, emotional engagement on Site C, which is sensitive to content length and mode changes, can be explained through the segmentation and cognitive load management principles from CTML (Mayer, 2021): prolonged multimedia presentation without mode variation increases the likelihood of cognitive fatigue and decreases sustained attention.

The negative case at Site B, where reduced focus followed from insufficient rather than excessive exposure, complicates a direct application of the CTML segmentation principle. This pair of contrasting negative cases, when read in conjunction with Site C's fatigue-from-overexposure case, supports a non-linear relationship between media exposure intensity and engagement quality, rather than the simple positive relationship suggested by the segmentation principle alone. Given the two negative cases in an overall nine-participant qualitative design, these observations should be interpreted as hypothesis-generating, rather than as evidence for the existence of an inverted-U curve or for a particular optimum point of exposure; the question of whether an inverted-U curve exists, and of where the optimum exposure point may lie, will remain an issue for future research that will examine the exposure-engagement relationship more systematically.

Grassini (2023) reports that engagement in generative AI-based learning is strongly shaped by the fit between tool affordances and individual learning preferences, consistent with the finding that engagement forms differed across sites despite identical media. Grothaus (2023) similarly finds that variation in student engagement is shaped by instructional practice, institutional arrangement, and sociocultural context, indicating that engagement patterns cannot be explained by technology alone. Fredricks et al. (2004) provide a conceptual basis for the three engagement forms identified here with their cognitive/affective/behavioral engagement typology.

The main conceptual contribution of this theme is a typology of institutional engagement orientations: competitive-performative (public junior high schools), affective-relational (pesantren), and reflective-metacognitive (private junior high schools), and a contextual framework explaining why identical media stimuli produced qualitatively different engagement manifestations. Practically, competition-based worksheets and

scoreboards may suit public junior high schools; post-video group discussion may suit pesantren; and modular 5-8-minute videos with modality transitions every 2-3 minutes may suit private junior high schools.

Understanding (CTML). These findings support CTML's core principles (Mayer, 2021), particularly the dual-channel assumption that verbal and visual information are processed through independent but complementary cognitive channels when coordinated (the multimedia principle) and appropriately matched to modality (the modality principle). P2's description of synergy among images, sound, and animation empirically illustrates the multimedia principle in a junior-high EFL context.

The negative case concerning audio quality offers an important, if unintended, confirmation of CTML's coherence principle. Mayer (2021) argues that extraneous elements can overload working memory and disrupt the processing of essential information, which is consistent with P1's complaint that background music obscured vocal clarity. The articulation problems P3 identified are similarly relevant to the modality principle: when the auditory channel is unclear, verbal-auditory working memory becomes overloaded, forcing greater reliance on the visual-textual channel, a compensation not always available without subtitles.

This is a technical failure in AI audio production, not of instructional sequencing or script design, and is a type of failure not yet well formulated in CTML, and prior to the advent of generative AI content. Theoretically, prosodic imbalance between AI narration and background music can significantly increase extraneous cognitive load (Lim et al., 2023). Notably, the findings in this study reveal a boundary condition for the positive effects of the cognitive benefits of multimedia design that has not been discussed in previous research, such as Kim (2024) and Fitria (2024), who reported mostly positive results for AI narration clarity because they used closed-survey instruments that do not explicitly allow for spontaneous technical evaluation. The methodological comparison demonstrates the added value of semi-structured interviews for uncovering critical evaluations that may be overlooked in closed-ended questionnaires.

The implications of these are: (1) making it a norm for using EFL AI videos to have subtitles; (2) during the production process, attention should be paid to the volume of the narration and the music so that the speech is clearly audible; (3) phonetic annotation for target vocabulary should be made; and (4) replay features with segment-based navigation should be incorporated to facilitate students' active remediation strategies as observed at Site C.

Autonomy (Benson) and institutional epistemology. According to Benson (2011), the Learner Autonomy Framework suggests that autonomy is not just a skill with two sides, but a construct with multiple dimensions that develops through the interaction between technical dimensions (mastery of self-directed learning strategies), psychological dimensions (the desire to control the learning process), and socio-political dimensions (the

structural space and recognition provided by the surrounding environment). The three levels of continuum seen in this study—aspirational autonomy (Site A), conditional autonomy (Site C), and structurally heteronomous (Site B)—directly reflect this multidimensional complexity.

The explicit, consistent rejection of independent media use by all three Site B participants is best explained through Benson's political-social dimension rather than through technical or psychological dimensions alone. This pattern is consistent with the epistemology of pesantren education at Site B, historically rooted in the tradition of *sanad* and *ijazah*; participants' rejection of independent media use may reflect a broader institutional orientation in which technology-based self-directed learning is seen as less epistemically legitimate than teacher-delivered knowledge, although this study did not directly ask participants about their views on *sanad* or *ijazah*, so this interpretation should be read as consistent with, rather than confirmed by, the interview data. This reading aligns with Benson's emphasis on the political-critical dimension of autonomy, which is a question that precedes questions of method or media and has an affinity with Bourdieu's (1977) concept of epistemic habitus. However, it is operationalized here at the institutional rather than individual level, as institutional epistemology.

These findings align with Reynolds et al. (2022), who report that students with adequate access to learning technology consistently show increased learning independence across institutional contexts; the Site B case indicates that technology availability is insufficient when institutional ideology does not legitimize technology-based learning autonomy as a valid practice. The results align with Chiu et al. (2023), who state that it is not only the technological affordances that are important in determining the effectiveness of AI-assisted learning, but also teacher support and pedagogical strategies that promote self-directed learning; coupled with other evidence that autonomy in AI-assisted learning is contingent upon contextual and instructional factors, this implies that learning culture and teacher expectations play a significant role in how much learner autonomy AI-based media can provide.

Site C's transitional nature resonates with the notion of scaffolded autonomy in the current literature on second-language autonomy. This aligns with Kim (2024), who found that EFL adolescents rarely exhibit full autonomy but do so by traversing stages through the mediation of social support. In reality, in teacher-authority-oriented epistemologies of institutions, EdTech interventions designed for autonomy might have to be packaged as teacher-guided extension tools, used collaboratively or under teacher guidance to make them legitimate in the institutional context.

The CM-TAEM Model: Institutional Moderation as an Integrative Framework

The convergence-divergence pattern recurring across the RQ1 and RQ2 findings motivates the proposed Contextually-Mediated Technology Acceptance and Engagement

Model (CM-TAEM), which connects TAM, SDT, CTML, and the Learner Autonomy Framework through two layers of contextual moderators. **Layer 1, Institutional Culture:** institutional epistemology, relational orientation (teacher-centered versus peer-centered), and institutional engagement orientation (competitive, affective, or reflective). **Layer 2, Infrastructure Equity:** device availability, connection stability, and institutional device-access policy.

These two layers of moderators determine the interaction between media inputs (AI video and Canva worksheets) and four categories of outcomes: technology acceptance (TAM), learning motivation (SDT), student engagement (SDT + CTML), and learning autonomy (Benson), where understanding EFL material becomes an outcome influenced by all four categories and also affected by AI audio quality.

CM-TAEM should therefore be understood as a theoretically derived and contextually grounded framework, an emergent conceptual model built from this study's qualitative findings, rather than a quantitatively validated model. The four emergent contextual dimensions, extrinsic PEU, relational channel specificity, institutional epistemology, and institutional engagement orientation, are analytic categories drawn from a multi-site qualitative dataset with nine participants. They are presented here as context-sensitive conceptual dimensions, not as validated constructs, and would need to be operationalized as quantitative instruments before formal testing (Table 5).

Table 5
Operationalization of Emergent Contextual Dimensions Informing the Proposed CM-TAEM Framework

Emergent Dimension	Operational Definition	Data Basis (Site)	Measurement Implications
Extrinsic PEU	Perceived ease of use conditioned by infrastructural factors beyond users' control (signal quality, device policy, equipment availability), distinct from intrinsic PEU inherent in interface design.	Site B: signal barriers and pesantren device policies. Site C: dependence on school tablets and internet connectivity	Separate Likert scale for infrastructural barriers, distinct from interface perception; adaptable from UTAUT facilitating conditions (Venkatesh et al., 2003)
Relational Channel Specificity	The tendency to fulfill relatedness needs (SDT) through specific channels determined by the institution's relational culture, which technological channels cannot substitute	Site B: negative case, all three participants consistently rejected the substitution of the teacher by digital media as a source of relatedness	Relatedness-channel-preference instrument (human vs. technology); in-depth interviews in teacher-centric contexts; relevant for pesantren institutions and collectivist cultures
Institutional Epistemology	Institutional beliefs about the locus of	Site B: <i>sanad</i> and <i>ijazah</i> traditions of pesantren	Instrument measuring teacher and student

	legitimate knowledge transmission, operating as a dominant filter over the perceived legitimacy of technology-based self-directed learning, are rooted in institutional pedagogical tradition.	education. Site A: academic-achievement orientation. Site C: curricular pragmatism	perceptions of knowledge authority; institutional policy-document analysis; principal interviews; affinity with Bourdieu's (1977) epistemic habitus.
Institutional Engagement Orientation	A typology of learning-engagement tendencies produced by institutional logic: (a) competitive-performative (public junior high school); (b) affective-relational (pesantren); (c) reflective-metacognitive (private junior high school)	Distinct engagement patterns across all three sites despite identical media stimuli; confirmed by classroom observation and teacher reflection notes	Structured observation instrument based on the three-orientation typology; validation through a larger-scale mixed-methods survey-observation design

Read together, the site-specific patterns extend each of the four guiding theories distinctly: TAM gains an infrastructure-facilitating-conditions dimension operating independently of interface design; SDT's relatedness construct gains motivational weight from teacher-centered vertical relationships, not only horizontal peer relationships; CTML gains AI audio-engineering quality as a previously unarticulated boundary condition; and the political-social dimension of the Learner Autonomy Framework becomes sharper through a three-level autonomy pattern and through institutional epistemology as a filter on the perceived legitimacy of technology-based self-directed learning. It is precisely the convergence of these institution-specific extensions across four otherwise separate theories that motivates treating institutional culture and infrastructure equity as structural moderators rather than background variables, the premise on which CM-TAEM is built.

At the policy level, the infrastructure gaps observed across sites suggest that digital equity is a structural prerequisite policymakers need to address before the potential of AI-based learning technology can be realized evenly across Indonesian school contexts, without widening gaps in English learning outcomes between school types.

CONCLUSION

Based on the results of this study, the videos that were made with AI and the worksheets that Canva made were appreciated by the seventh-grade EFL students in the three cases of institutions in this study. The means of this value, however, varied with context. The perceived usefulness was quite uniform, whereas the perceived ease of use had both interface and infrastructure dimensions, with infrastructure-related constraints most apparent in the pesantren. These were institutionally different patterns in relation to the variables of Motivation, Engagement, and Understanding, and the most different patterns

were found in the variable of Learner Autonomy, where the public school was aspirational, the private school conditional, and the pesantren structurally constrained.

Based on these patterns, a Contextually Mediated Technology Acceptance and Engagement Model (CM-TAEM) was proposed that combines institutional culture and institutional infrastructure equity with the dimensions of TAM, CTML, SDT, and the Learner Autonomy Framework, resulting in four emergent dimensions: extrinsic PEU, relational channel specificity, institutional epistemology, and institutional engagement orientation. CM-TAEM remains theory-based and context-sensitive and has not been empirically or quantitatively validated. The main message is that AI's contribution to EFL learning needs to be interpreted within its institutional context, not simply in terms of the technology.

The limitations are the small purposive sample, the cross-sectional design across three sites, and the reliance on self-report and observation. Future studies should validate CM-TAEM quantitatively or using mixed methods, investigate learner autonomy in a longitudinal study, and test specific interventions (low-bandwidth video delivery, subtitles, and guided autonomy scaffolding).

ACKNOWLEDGMENT

The authors would like to express their deep thanks to the students, teachers, and school staff at the three junior high schools for their participation and support in this research. We also appreciate the role of the independent interviewers and colleagues who helped with the debriefing process, including data collection and analysis. The authors take full responsibility for all the content, accuracy, and integrity of this article.

DECLARATION OF AI USE

The authors recognize the potential for limited use of artificial intelligence (AI) technology to assist with the initial draft and to verify grammar, language clarity, and sentence structure. All AI-generated content has been checked, validated, and edited by the authors. Research data were not created using AI, and AI was not used for analysis, interpretation, or to derive research conclusions; the researchers themselves made all analytical and interpretive decisions.

REFERENCES

- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50(1), 179-211. [10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Al-Khasawneh, F., Huwari, I., Alqaryouti, M., Alruzzi, K., & Rababah, L. (2024). Factors Affecting Learner Autonomy in the Context of English Language Learning. *Cakrawala Pendidikan*, 43(1), 140–153. <https://doi.org/10.21831/cp.v43i1.61587>
- Amin, A.-R., E. (2024). EFL Students' Perception of Using AI Text-to-Speech Apps in Learning Pronunciation. *SSRN Electronic Journal*, 8984, 887–903. <https://doi.org/10.2139/ssrn.4746800>

- Benson, P. (2011). *Teaching and Researching Autonomy* (2nd ed.). Pearson Education.
- Bourdieu, P. (1977). Outline of a Theory of Practice. *Translated by*. Cambridge University Press.
- Chiu, T. K. F., Moorhouse, B. L., Chai, C. S., & Ismailov, M. (2023). Teacher Support and Student Motivation to Learn with Artificial Intelligence (AI) Based Chatbot. *Interactive Learning Environments*, 32(7), 3240–3256. <https://doi.org/10.1080/10494820.2023.2172044>
- Creswell, J. W., & Cheryl N. Poth. (2018). Qualitative Inquiry & Research Design. In *Journal of Physics A: Mathematical and Theoretical* (FOURTH EDI, Vol. 44, Number 8). SAGE Publications.
- Creswell, J. W., & Poth, C. N. (2018). *Qualitative Inquiry and Research Design: Choosing Among Five Approaches* (4th ed.). SAGE Publications.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly: Management Information Systems*, 13(3), 319–339. <https://doi.org/10.2307/249008>
- 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
- Dörnyei, Z., & Ushioda, E. (2021). *Teaching and researching motivation: The third edition*. Routledge.
- Fitria, T. N. (2024). Using Canva in Creating and Designing E-Module for English Language Teaching. *SAGA: Journal of English Language Teaching and Applied Linguistics*, 5(2), 111–122. <https://doi.org/10.21460/saga.2024.52.194>
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School Engagement: Potential of the Concept, State of the Evidence. In *Review of Educational Research* (Vol. 74, Number 1, pp. 59–109). <https://doi.org/10.3102/00346543074001059>
- Fujii, A. (2024). Exploring Autonomy Support and Learning Preference in Higher Education: Introducing a Flexible and Personalized Learning Environment With Technology. *Discover Education*, 3(1) 1–16. <https://doi.org/10.1007/s44217-024-00111-z>
- Godwin-Jones, R. (2021). Big Data and Language Learning: Opportunities and Challenges. *Language Learning and Technology*, 25(1), 4–19. <https://doi.org/10.64152/10125/44747>
- Grassini, S. (2023). Shaping the Future of Education: Exploring the Potential and Consequences of AI and ChatGPT in Educational Settings. *Education Sciences*, 13(7), 2-13, <https://doi.org/10.3390/educsci13070692>
- Grothaus, C. (2023). Engagement in Online Learning Among Thai and German students: The role of Classmates, Instructors, and Technology across Country Contexts. *Online Learning Journal*, 27(3), 188–208. <https://doi.org/10.24059/olj.v27i3.3413>
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations* (2nd edition). Sage Publications.
- Kemendikbudristek. (2022). *Panduan Pembelajaran dan Asesmen: Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah*. Jakarta: Badan Standar, kurikulum, dan Asesmen Pendidikan .

- Kim, R. (2024). Comparative Analysis of AI-Generated and Human-Made Instructional Videos: Effects on Learning English Caused-Motion Constructions by Korean High School EFL Learners. *Multimedia-Assisted Language Learning*, 27(4). 180–199, <https://doi.org/10.15702/mall.2024.27.4.179>
- Lim, W. M., Gunasekara, A., Pallant, J. L., Pallant, J. I., & Pechenkina, E. (2023). Generative AI and the Future of Education: Ragnarök or Reformation? A Paradoxical Perspective From Management Educators. *International Journal of Management Education*, 21(2), 100790. <https://doi.org/10.1016/j.ijme.2023.100790>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic Inquiry*. SAGE Publications, 438–439.
- Ly, D. T. (2024). EFL Students' Perceptions on the Use of Liveworksheets in Self-study. *Proceedings of International Conference on Research in Education and Science*, 10(1), 838–849. www.istes.org
- Malterud, K., Siersma, D., V., & Guassora, D. (2016). Sample Size in Qualitative Interview Studies: Guided by Information Power. *Qualitative Health Research*, 26(13), 1–8. <https://doi.org/https://doi.org/10.1177/1049732315617444>
- Matthew B. Miles, A. Michael Huberman, J. S. (2014). *Qualitative Data Analysis: A Methods Sourcebook*. In SAGE Publications (Edition 3). 978-1-4522-5787-7.
- Mayer, R. E. (2021). *Multimedia Learning* (3rd ed.). Cambridge University Press.
- Merriam, S. B., & Tisdell, E. J. (2016). *Qualitative Research: A Guide to Design and Implementation*. Jossey-Bass, a Wiley Brand.
- Mutiaraningrum, I., & Nugroho, A. (2021). Smartphone-Based Mobile Assisted Language Learning Application in Higher Vocational Education in Indonesia. *JEES (Journal of English Educators Society)*, 6(1) 26-34. <https://doi.org/10.21070/jees.v6i1.793>
- Pekrun, R., W. Marsh, H., J. Elliot, A., Stockinger, K., P. Perry, R., Vogl, E., Goetz, T., A. P. van Tilburg, W., & Lüdtke, Oliver Walter P. Vispoel1, W. (2026). A Three-Dimensional Taxonomy of Achievement Emotion. *Journal of Personality and Social Psychology: Personality Processes and Individual Differences*, 124(1), 145–178. <https://doi.org/https://doi.org/10.1037/pspp0000448>
- Phi, The, T., & Khanh, Gia, N. (2024). EFL Learners' Perceptions of Conversational Videos Regarding Classroom Engagement at a Language Center in Can Tho City. *International Journal of TESOL & Education*, 4(3). 131- 143. <https://doi.org/10.54855/ijte.24438>
- Purwasi, L. A., Fitriyana, N., Jayati, R. D., Fitriani, L., & Irawan, B. P. (2023). *Developing Interactive E-Modules Through Canva App* (Vol. 1). Atlantis Press SARL. https://doi.org/10.2991/978-2-38476-108-1_32
- Patton, M. Q. (2015). *Qualitative Research and Evaluation Methods* (4th ed.). SAGE Publications.
- Reynolds, B. L., Yang, H., Song, H., & Liu, Z. (2022). Language Teaching with Video-Based Technologies: Creativity and CALL Teacher Education. *Frontiers in Psychology*.
- Savitri, F. A., & Ratri, D. P. (2025). Canva as a Creative Digital Learning Media for English Learning: A Student Perception Study. *Didaktika : Jurnal Kependidikan*, 14(2), 1811–1820. <https://doi.org/10.58230/27454312.2184>
- Stake, R. E. (1995). *The Art of Case Study Research*. SAGE Publications.
- UNESCO. (2023). *Global Education Monitoring Report 2023: Technology in Education: A Tool on Whose Terms?*

- Vansteenkiste, M., Ryan, R. M., & Soenens, B. (2020). Basic Psychological Need Theory: Advancements, Critical Themes, and Future Directions. In *Motivation and Emotion* (Vol. 44, Number 1). Springer US. <https://doi.org/10.1007/s11031-019-09818-1>
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. In *MIS Quarterly* (Vol. 27, Number 3). <https://www.jstor.org/stable/30036540>
- Wei, L. (2023). Artificial Intelligence in Language Instruction: Impact on English Learning Achievement, L2 Motivation, and Self-Regulated Learning. *Frontiers in Psychology*, 14(November), 1–14. <https://doi.org/10.3389/fpsyg.2023.1261955>
- Wei, Y. (2022). Toward Technology-Based Education and English as a Foreign Language Motivation: A Review of Literature. *Frontiers in Psychology*, 13. <https://doi.org/10.3389/fpsyg.2022.870540>
- Xue, L., Mahat, J., & Ghazali, N. (2026). Technology Acceptance Model in Artificial Intelligence in Education: A Meta-Analysis. *SAGE Open*, 16(1), 1–23. <https://doi.org/10.1177/21582440251409441>
- Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th Edition, Vol. 2). SAGE Publications.
- Zhang, Y., Lucas, M., Bem-haja, P., & Pedro, L. (2024). The Effect of Student Acceptance on Learning Outcomes: AI-Generated Short Videos Versus Paper Materials. *Computers and Education: Artificial Intelligence*, 7(June), 100286. <https://doi.org/10.1016/j.caeai.2024.100286>