



Castledown

 OPEN ACCESS

Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

Technology in Language Teaching & Learning, 8, 104691 (2026)

<https://doi.org/10.29140/tltl.2026.104691>

Learning Outcomes and Motivation in a Gamified, Lightly Personalised Virtual Learning Environment for Secondary English



MYRIAM TATIANA VELARDE OROZCO^a 

BÁRBARA LUISA DE BENITO CROSETTI^b 

^a*Doctoral Programme in Educational Technology,
Doctoral School, University of the
Balearic Islands, Spain
myriam-tatiana.velarde@uib.cat*

^b*Department of Applied Pedagogy and Educational
Psychology, Faculty of Education, University of the
Balearic Islands, Spain
barbara.debenito@uib.es*

Abstract

This study reports a field evaluation of VLEPIC, a gamified and lightly personalised virtual learning environment developed to support English learning in a public secondary school in Riobamba, Ecuador. As part of the second iteration of a design-based research project, the study explored learning outcomes, learner motivation, and the relationship between them in 131 students aged 12–15 from six intact classes. VLEPIC was used over one school term as teacher-assigned practice outside regular class time, and the quantitative findings were complemented with learners' open-ended comments. Students showed measurable improvement from pretest to posttest, with mean scores increasing from 9.19 to 11.92 and a statistically significant Wilcoxon result with a large rank-based effect size. Motivation was positive across all four dimensions, with relevance and attention receiving the highest ratings. Learners frequently said they felt motivated by the platform's visual and interactive design, the feeling of moving forward through missions and rewards, and the guidance provided while using it. At the same time, some reported reduced motivation when they had to record tasks, deal with text-heavy activities, face comprehension difficulties, cope with early navigation issues, or feel pressure related to scores. Only satisfaction was significantly associated with learning gain at the correlational level, but none of the motivational dimensions remained significant once baseline performance was taken into account. Overall, the findings suggest that VLEPIC was associated with a positive learning experience and measurable short-term progress in this Ecuadorian public-school context, while also indicating that baseline achievement remained more consequential for posttest performance than motivational ratings.

Copyright: © 2026 Myriam Tatiana Velarde Orozco & Bárbara Luisa de Benito Crosetti. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Keywords: English as a foreign language, gamification, learner motivation, learning outcomes, personalised learning, secondary education, virtual learning environment.

Introduction

Schools are increasingly adopting digital tools to support learning (Forsström et al., 2025). In digital language learning (LL) environments, an important question is whether these tools only offer usable and engaging experiences or also contribute to meaningful learning outcomes (Forsström et al., 2025; Lu et al., 2025). Although usability and user experience (UX) are relevant for understanding how learners interact with a system, the evaluation of learning technologies also requires attention to pedagogical outcomes beyond technological usability (Lu et al., 2025). For this reason, classroom-based studies on technology in LL should examine evidence of learning alongside learners' perceptions of the environment (Chan & Lo, 2024; Shen et al., 2024).

This need is especially relevant for virtual learning environments (VLEs) that add gamification and personalisation. These features are often introduced to support sustained practice and learner engagement over time, yet their presence does not guarantee better learning outcomes in real classrooms (Chan & Lo, 2024; du Plooy et al., 2024). In addition, motivation is often discussed as one reason why some learners persist and benefit more than others in technology-enhanced learning (TEL), including online LL (Shen et al., 2024). A practical way to describe motivation at the level of instructional design is Keller's ARCS framework (attention, relevance, confidence, and satisfaction) (Fang et al., 2024; Keller, 2010). In this paper, these four dimensions are used to describe learners' motivational profile and examine how it relates to learning outcomes.

This study contributes to research on technology-supported English learning in three main ways. First, it offers evidence from a real public-school context in Ecuador, where access to technology is not always equal for all students and where school-based studies on gamified VLEs are still scarce. Second, it focuses on learning outcomes along with learner motivation, so the focus is not only on whether students liked the platform, but also on whether that positive experience was linked to short-term learning progress. Third, it evaluates a lightly personalised VLE under real implementation conditions, documenting short-term learning outcomes and motivational responses while also examining the role of students' baseline achievement in posttest performance. The study seeks to answer three research questions (RQs):

RQ1: Do learners show measurable gains from pretest to posttest after using VLEPIC?

RQ2: What motivational profile do learners report across attention, relevance, confidence, and satisfaction?

RQ3: Are these motivational indicators associated with learning gains?

Literature Review

Motivation in Technology-mediated Language Learning

In contexts specific to TEL, motivation is frequently considered situational, as it can vary with the task, the materials, and the conditions of use (Keller, 2010). Keller's ARCS model, for example, is often used as a theoretical basis for exploring how instructional design affects engagement and persistence (Keller, 2010). A recent systematic review indicates that ARCS has been applied across a growing range of educational and research contexts, including TEL; yet differences exist between studies regarding how they apply the framework to practice (Fang et al., 2024).

Within the context of ARCS theory, Attention promotes motivation in learning activities by maintaining learners' interest; Relevance addresses whether they perceive a clear connection to their goals or needs; Confidence corresponds to expectations of success and the feeling that progress is realistic; while Satisfaction focuses on post-task positive feelings and perceived value (Keller, 2010). In educational research, ARCS is often operationalised via ARCS-aligned self-report measures completed by learners post-activity. These are usually interpreted in relation to the specific learning context and the way the activity is delivered (Fang et al., 2024).

Motivation is also relevant when examining learning outcomes, although the relationship is not necessarily direct or uniform. In online LL, Shen et al. (2024) found that learner motivation mediated the relationship between gamification integration and LL achievement, suggesting that motivational processes may help explain how learners respond to gamified environments. Other LL research has similarly reported concurrent improvements in language achievement and affective outcomes such as foreign language enjoyment and ideal L2 self, following digital gamified learning (Liu et al., 2024). These findings support examining motivational responses alongside learning outcomes while avoiding the assumption that positive motivational ratings necessarily translate into greater achievement.

Gamification in Technology-supported English Learning

Gamification is the application of game elements to learning contexts that would not otherwise use games (Deterding et al., 2011). When integrated into instructional practice, it may help learners stay engaged over time. In the context of English learning, reviews specifically on English as a Foreign Language (EFL) and English as a Second Language (ESL) suggest that gamification tends to correlate positively with both language-related and affective outcomes, although findings diverge considerably depending on the contexts and methodologies used (Chan & Lo, 2024; Zhang & Hasim, 2023). A systematic review in CALL similarly outlines generally positive effects for learning experiences and outcomes in gamified English learning yet finds that studies often fail to untangle which specific game elements relate to outcomes (Dehghanzadeh et al., 2021). Recent synthesis work covering 2020–2024 further articulates benefits alongside deployment complexities and methodological variability (Arip & Hashim, 2024). These reviews support examining learning and motivational outcomes concurrently, while also indicating that both should be interpreted in relation to the context and conditions of implementation.

Overall, some mechanics appear more consistently than others across ELT-oriented VLEs and have been linked to motivational outcomes. More specifically, our systematic review revealed that challenge-based mechanics, points, and rewards were frequently related to increased motivation and engagement, particularly when adaptive feedback was available (Velarde Orozco & de Benito Crosetti, 2025). Empirical evidence has similarly shown that specific gamification mechanics can shape patterns of engagement and learning outcomes in technology-supported EFL practice (Lafleur, 2024). Beyond direct associations with learning and affective outcomes, recent evidence suggests that motivation may function as a mechanism through which gamification is related to LL achievement (Shen et al., 2024).

At the same time, it is well-documented in the literature that gamification is not a “one-size-fits-all” concept. The difference lies in the way game elements are integrated with tasks, feedback, and curriculum demands. What works in one context does not necessarily work in another. Our systematic review emphasised the importance of alignment with curricular objectives and resource constraints (Velarde Orozco & de Benito Crosetti, 2025), while Simsek and Karakus Yilmaz (2025) examined differences across subject areas, sample characteristics, implementation forms, and technologies. Geographic and linguistic biases in the literature, together with context-sensitive and sometimes mixed findings, caution against generalising findings derived mainly from better-represented settings (Chan

& Lo, 2024; Shang, 2025). This limitation is particularly relevant to the present study, which examines learning outcomes and motivational responses in an Ecuadorian public-school context and interprets them in relation to the conditions under which VLEPIC was implemented.

Personalisation in Technology-supported English Learning

In TEL, personalisation is a catch-all term that may overlap with similar descriptors, including adaptive, individualised, or customised learning (Bernacki et al., 2021; Shemshack & Spector, 2020). In this literature, personalisation most often refers to adapting content, sequence, pacing, difficulty, or support based on learner characteristics, progress, and/or learner choices (Bernacki et al., 2021; du Plooy et al., 2024). Considering that definitions and implementations differ, it is usually more informative to provide a descriptive account of personalisation in design terms, specifying what changes, when it changes, and who controls it, instead of using a label as a proxy for a specific mechanism (Bernacki et al., 2021; Shemshack & Spector, 2020).

Evidence syntheses show that personalised learning (PL) encompasses diverse designs that often combine multiple components, making it difficult to attribute outcomes to individual features without clear reporting of how personalisation is implemented (Bernacki et al., 2021). In higher education, personalised adaptive learning has shown generally positive academic outcomes, although effects on engagement are less consistent and implementation may be constrained by technological and time-related demands (du Plooy et al., 2024). In secondary education, teachers have also reported practical complexities associated with implementing PL (Gunawardena et al., 2024).

Adaptive and personalised approaches differ in the degree and mechanism through which learning is adjusted. Reviews of PL emphasise the diversity of definitions and implementations and the need to specify the particular mechanisms used in each intervention (Bernacki et al., 2021; Shemshack & Spector, 2020). In technology-assisted language learning, adaptive systems may modify content, feedback, or learning pathways according to learners' knowledge, performance, or other characteristics (Kaur et al., 2023). At the same time, research in secondary education shows that implementing PL can involve considerable practical complexity from the teacher perspective (Gunawardena et al., 2024). These distinctions are relevant to the present study because VLEPIC used a lighter form of personalisation based on diagnostic placement and preference-based support within a common learning structure, without continuous algorithmic adaptation.

Methods

Research Design

This study represents the second design iteration within a design-based research (DBR) programme. DBR aligns closely with educational technology research, as it integrates the design of an intervention with empirical inquiry in real-world settings and uses evidence from implementation to refine the intervention iteratively (Anderson & Shattuck, 2012; Design-Based Research Collective, 2003; McKenney & Reeves, 2018). The evaluation discussed here assesses learning outcomes and learner motivation within use cases aligned with the school curriculum.

The present iteration was informed by the preliminary design principles generated in the first VLEPIC evaluation (Velarde Orozco & de Benito Crosetti, 2026a). Findings from that iteration led to a stronger emphasis on structural and visual coherence, clearer progress and navigation feedback, differentiated scaffolding, mobile usability, reliable task submission, and visible progression and rewards. These refinements shaped the version of VLEPIC evaluated in the present study and illustrate the iterative link between empirical evidence and design decisions within the DBR programme.

This study employs a predominantly quantitative design supplemented by descriptive qualitative feedback. Combining quantitative and qualitative evidence can support a more comprehensive understanding of complex educational phenomena and innovations (Creswell, 2022; Creswell & Plano Clark, 2017; Ramírez-Montoya & Lugo-Ocando, 2020; Tashakkori & Teddlie, 2010). In the present study, the quantitative dimension gauges changes in performance while describing patterns across motivational ratings; meanwhile, responses to a single open-ended survey item are used descriptively to help interpret observed trends.

The primary quantitative approach for assessing learning outcomes is a one-group pretest-posttest design. This design can track change over time; however, it does not allow for strong causal inferences, as alternative explanations for such change cannot be eliminated (Campbell & Stanley, 1963; Shadish et al., 2002). This refers to typical threats such as testing effects, history, and maturation (Campbell & Stanley, 1963; Shadish et al., 2002). Consequently, findings are reported and understood as evidence of differences consistent with the intervention in the defined contexts of use, as opposed to proof of effect (Bierer et al., 2025; Shadish et al., 2002). To support transparent interpretation, the study reports effect sizes and focuses on practical significance instead of solely relying on statistical significance (Shadish et al., 2002).

Participants

This study was conducted at Escuela de Educación Básica “San Felipe Neri”, a public school located in Riobamba, Ecuador. The study focused on lower secondary education and involved 131 EFL students aged 12–15 years, recruited from six intact classes: two eighth-grade, two ninth-grade, and two tenth-grade groups. A significant proportion of the student population comes from indigenous and low-income households. The school context is resource-constrained, with sparse technology infrastructure and varying levels of student access to devices and the internet.

Ethical Considerations

The study was conducted in accordance with the Declaration of Helsinki (World Medical Association, 2024) and approved by the Research Ethics Committee of the University of the Balearic Islands (protocol code 051CER25 and date of approval 18 September 2025). Parents or guardians signed informed consent forms, and students provided assent. Participation was voluntary, and students were free to withdraw from participation at any time without academic penalty.

All datasets were anonymised prior to any analysis or reporting to protect privacy. Quantitative results are reported in aggregate, while illustrative quotations are presented using anonymised participant codes to minimise the potential for individual identification. Data management was undertaken using an approach consistent with confidentiality, equity, and non-discriminatory participation, as well as the protection of the rights of participants, following international guidance for educational research involving human beings (British Educational Research Association, 2024).

Intervention

During the intervention period, VLEPIC was used over one school term (one trimester) as teacher-assigned practice outside regular class time. The content of the platform was aligned with the Ecuadorian national EFL curriculum (Ministerio de Educación del Ecuador, 2016). The platform was organised into four expeditions: Starter Base (SB), Novice Trail (NT), Explorer’s Path (EP), and Mastery Quest (MQ). Each expedition included six quests, except for SB, which included four, and each quest was structured around five missions. Progression through the platform was sequential and blocked, so students had to complete each stage before accessing the next one. The student dashboard

provided learners with a clear entry point to the main areas of VLEPIC, including the diagnostic test, announcements, chat room, expeditions, and so on, as illustrated in Figure 1.

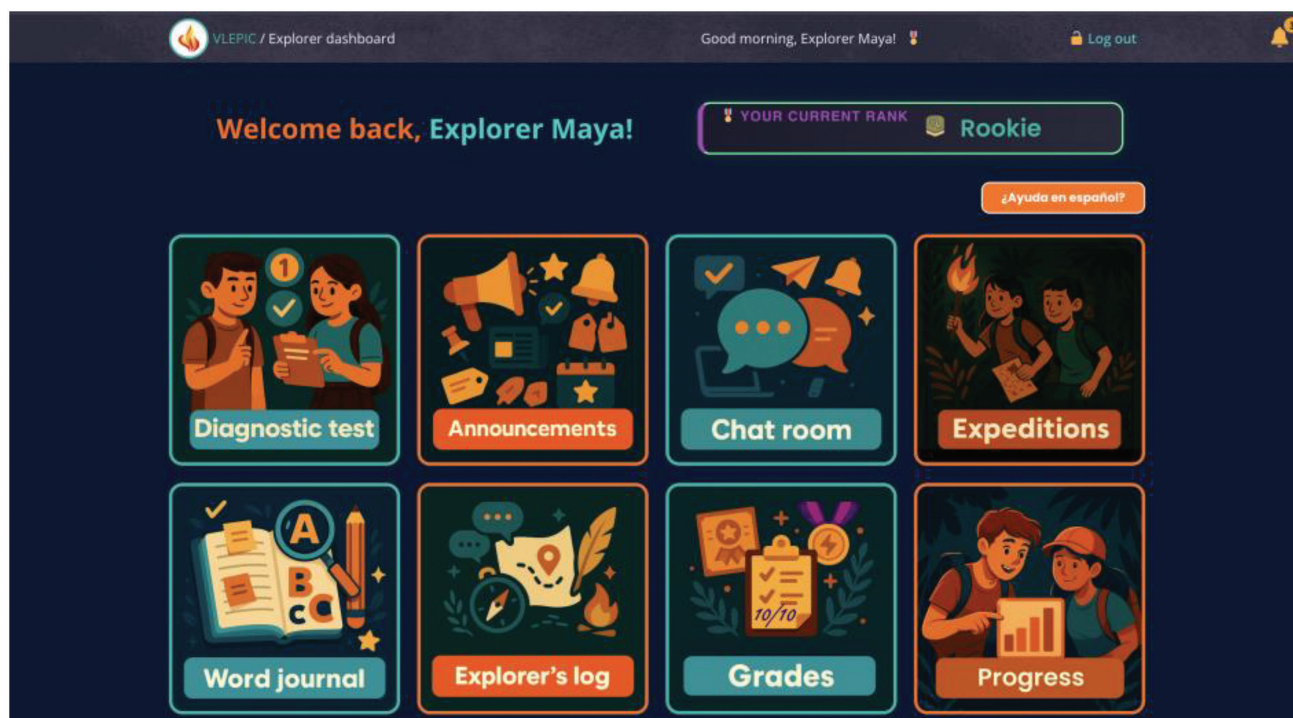


Figure 1 *VLEPIC student dashboard used during the intervention.*

This structure formed the main gamified progression system of VLEPIC. Missions were unlocked through badges, quests were associated with shields, and students advanced through a rank system linked to each expedition, from Rookie in SB to Master at the highest level. In addition, students earned experience points (XP), which could later be exchanged for real classroom rewards.

During the term covered in this study, students were expected to complete two quests (units of learning), each comprising five missions (structured lessons), within their assigned pathway. Platform logs indicated 856 successful login events across the 131 registered learners and 1,317 task submissions from 125 students. A login event was defined as each successful authentication recorded by the platform; therefore, the same learner could generate multiple login events during the term. Students used the platform for specific tasks at different times during the term, and participation varied across the mission sequence. The number of learners submitting at least one task increased from 79 in Mission 1 to 97 in Mission 2 and 100 in Mission 3, before declining to 95 in Mission 4 and 85 in Mission 5 (Velarde Orozco & de Benito Crosetti, 2026b). Because the available logs were analysed at cohort level, they did not permit complete reconstruction of individual learning trajectories or exact quest completion for each learner. Accordingly, exposure to VLEPIC was not fully uniform, and the study should be interpreted as an evaluation conducted under authentic implementation conditions, with variable learner exposure and without a tightly controlled dosage structure.

Regular classroom instruction continued throughout the intervention, with VLEPIC functioning as supplementary asynchronous practice alongside classroom teaching. The first author coordinated the implementation of VLEPIC across the six participating classes, while regular classroom instruction was delivered by the English teachers responsible for each class. All groups followed the Ecuadorian

national EFL curriculum, while VLEPIC activities were assigned according to learners' diagnostic placement and therefore differed in level and content across pathways. The platform's interaction structure and progression logic remained consistent throughout the implementation (Velarde Orozco & de Benito Crosetti, 2026b). Nevertheless, regular classroom instruction was not experimentally standardised across intact classes, and possible class-level differences in instructional experience cannot be separated from the observed pretest–posttest change.

Before starting the intervention, students completed a diagnostic placement test within VLEPIC. This test was organised into three sequential blocks corresponding to SB, NT, and EP. Students began in SB and progressed to the next block only if they achieved the required threshold in the previous one, set at 80% correct responses. If they did not reach it, the test stopped at that point and the student was placed accordingly. Students who successfully completed all three blocks could enter MQ, which functioned as the highest starting level in the platform. The purpose of this test was instructional placement within VLEPIC, not outcome measurement.

Personalisation in VLEPIC was implemented as a light layer at the beginning of the experience. In addition to level placement through the diagnostic test, students could personalise their display name and choose learning preferences such as visual, auditory, kinaesthetic, read/write, and self-paced. Based on these choices, the platform provided short tips to support how they approached or reviewed the activities. No tips were provided for the self-paced option. In this study, personalisation was not designed as a fully adaptive system; it was implemented through level placement and preference-based support within the same platform.

Data Collection Instruments

Learning outcomes were assessed through separate researcher-developed pretest and posttest forms administered before and after the intervention period. These forms were distinct from the diagnostic placement test embedded in VLEPIC and were used only to assess change over time, not to assign students to a starting level. Each form comprised 30 items and included multiple-choice, completion, and selection tasks. Their content followed a progressive structure aligned with the learning sequence of VLEPIC, covering SB, NT, and EP. The two forms used different but matched items, with corresponding items addressing the same curricular content and language focus at each level. Their comparability was based on this structural and content matching; the forms were not statistically equated. The development of the forms was informed by recommendations concerning content representation in educational assessments designed for a specific purpose and context (American Educational Research Association et al., 2014; Sireci & Faulkner-Bond, 2014). The main characteristics of the instrument are presented in Table 1.

Table 1 *Structure and Content Coverage of the Learning Outcomes Instrument.*

Component	Section/level	Content focus	Learning focus	Item format
Pretest / Posttest	SB	Basic vocabulary and personal pronouns.	Recognition and basic use.	Multiple choice, completion, selection.
Pretest / Posttest	NT	Verb to be, present simple, and present progressive.	Controlled use of core grammatical structures.	Multiple choice, completion, selection.
Pretest / Posttest	EP	Past simple and past progressive.	Use of past forms in guided tasks.	Multiple choice, completion, selection.

(Continued)

Component	Section/level	Content focus	Learning focus	Item format
Overall structure	SB–EP	Progressive structure aligned with the learning sequence of VLEPIC.	Assessment of learning outcomes across three levels.	Mixed objective-format items.
Administration	Pretest / Posttest	Same overall structure; different but matched items across forms.	Measurement of change over time.	Parallel forms
Development process	Pretest / Posttest	Researcher-developed instrument.	Designed for the instructional context of the study.	Initial version piloted and revised; revised forms not re-piloted.

Before the main implementation, an initial version of the pretest-posttest instrument was piloted with 30 students to examine the clarity and appropriateness of item wording, level, and response format for the target group (American Educational Research Association et al., 2014; Sireci & Faulkner-Bond, 2014). The pilot results showed a mean score of 21.03 out of 30 ($SD = 3.29$), with an average percentage of correct answers of 70.11%. Item difficulty ranged from .43 to .87, with a mean difficulty of .70, suggesting that the initial version was generally accessible to the pilot group, but that some items required adjustment to better reflect the intended progression of difficulty. The KR-20 coefficient was modest (.47), and some items showed weak discrimination. Based on these results, specific refinements were made to increase item complexity where necessary, improve distractors, clarify wording, and ensure stronger alignment with the intended SB, NT, and EP levels, while the overall structure and progression of the test were retained. The development and revision process also involved a review by the school's English teachers using the Ecuadorian national EFL curriculum as the reference. The review considered whether the language content and intended level of each item corresponded to the curriculum objectives represented in the SB, NT, and EP levels. The revised forms were not re-piloted. Internal consistency was subsequently examined using the responses obtained from the main study sample. The final forms showed acceptable internal consistency in the pretest ($KR-20 = .731$) and good internal consistency in the posttest ($KR-20 = .851$). Each form included 10 items for each level (SB, NT, and EP), with one point assigned to each item. Students were allowed to leave items unanswered if they did not know how to respond. Unanswered items were scored as incorrect, and item scores were combined into an overall score ranging from 0 to 30. In this sense, the instrument was intended to provide context-specific evidence of short-term learning progress aligned with the instructional content of the intervention, not a standardised measure of general English proficiency.

Learner motivation was measured using the Instructional Materials Motivation Survey (IMMS), based on the ARCS model (Keller, 2010; Loorbach et al., 2015). The survey was administered in Spanish, following prior use of the instrument in educational technology research with Spanish-speaking learners (Di Serio et al., 2013). In addition to the closed-ended IMMS items, the survey included one open-response question asking students to identify which aspects of VLEPIC they felt supported or hindered their motivation to learn. This item was included to capture learner perceptions in their own words and to provide contextual information that could help to interpret the quantitative motivational profile. Given that the purpose of these comments was explanatory and not exploratory in depth, they were treated as a limited descriptive qualitative complement to the survey findings (Hsieh & Shannon, 2005; Sandelowski, 2000; Züll, 2016). The qualitative material was therefore used to identify recurrent motivating and demotivating aspects of the learning experience and to support interpretation of patterns across the ARCS dimensions.

Data Analysis

All quantitative analyses were performed in R (version 4.4.3). Data were screened for out-of-range values and missingness, with sample sizes reported for each analysis. The analyses were organised in relation to the RQs. Before analysis, the dataset was checked for completeness and consistency. Complete pretest, posttest, and IMMS scores were available for all 131 participants, so the same analytical sample was used for the pretest–posttest, correlation, and regression analyses. Gain scores were calculated as posttest minus pretest. Negatively worded IMMS items were reverse-scored before calculating the subscale and total mean scores; no other transformations were applied.

For RQ1, learning outcomes were examined through within-subject comparisons of pretest and posttest scores. When the distribution of difference scores met parametric assumptions, a paired-samples *t* test was used; otherwise, a Wilcoxon signed-rank test was applied. For parametric comparisons, Cohen's *d* for paired samples with 95% confidence intervals (CI) was reported (Lakens, 2013). For non-parametric comparisons, the effect size *r* associated with the Wilcoxon signed-rank test was also reported (Fiel Peres, 2026). In this study, meaningful gains were interpreted in terms of statistical significance, effect size, and the educational relevance of the observed change.

For RQ2, motivation was analysed across four subscales: Attention, Relevance, Confidence, and Satisfaction. Descriptive statistics were calculated for each subscale, including mean, standard deviation (SD), interquartile range (IQR), and 95% CI of the mean. Internal consistency was also estimated for each subscale and for the total IMMS score. Given the ordinal nature of the IMMS items and the limitations of relying exclusively on Cronbach's alpha, internal consistency was estimated using both ordinal alpha and McDonald's ordinal omega based on polychoric correlations (Dunn et al., 2014; Gadermann et al., 2012). The open-ended IMMS item was analysed through descriptive qualitative content coding. Responses were read repeatedly to gain familiarity with the data and then coded inductively, staying close to the wording used by participants. Initial codes were subsequently grouped into recurrent motivating and demotivating categories that summarised the aspects of VLEPIC most frequently mentioned by learners. After the initial coding, all responses were revisited to refine category boundaries and ensure consistency in category assignment. Considering that individual comments could refer to more than one aspect of the experience, a single response could be assigned to more than one category when appropriate. For frequency reporting, each explicit reference was counted as a separate mention; therefore, repeated references to the same category within one response contributed more than one mention. This analytic approach was consistent with the limited descriptive purpose of the qualitative component, which was intended to organise brief survey comments systematically and not to produce a deeper interpretive account (Hsieh & Shannon, 2005; Sandelowski, 2000; Züll, 2016). The resulting categories were used to contextualise the quantitative IMMS findings within the broader design of the study. Short illustrative quotations were retained to preserve learners' voices, and all quotations were translated from Spanish into English for reporting, with minor corrections for readability. The initial coding was conducted by the first author. The second author subsequently reviewed all 95 coded responses, together with the code structure, category definitions, and the fit between the coded responses and the resulting categories. Any ambiguities or overlaps were discussed until agreement was reached. Given the limited descriptive purpose of this qualitative component, no formal inter-rater reliability coefficient was calculated.

For RQ3, Pearson's correlations between motivational subscales and learning gains were used to examine bivariate associations between motivation and improvement from pretest to posttest. To further examine the adjusted association of each motivational dimension with posttest performance, four separate ordinary least squares (OLS) regression models were estimated, each including pretest

score as a covariate and one IMMS subscale as the motivational predictor. Subsequently, a full OLS regression model was estimated with posttest score as the outcome, the four motivational subscales as simultaneous predictors, and pretest score as a covariate (Vickers & Altman, 2001). This analysis was used to examine whether motivational ratings were associated with posttest performance once learners' baseline achievement and the other motivational dimensions were taken into account. Potential multicollinearity among the predictors in the full model was assessed using variance inflation factors (VIF) and tolerance values.

Results

Participant Profile

The final sample included secondary students from Grades 8 to 10, covering an age range from 12 to 15 years. Overall, the sample showed some variation in gender distribution across grades, with a higher representation of male students in the upper grades. Table 2 describes the demographic profile of the participants.

Table 2 *Participants' Profile*

Students	Female	Male	Age range	n
Grade 8	22	19	12–13	41
Grade 9	19	27	13–14	46
Grade 10	14	30	14–15	44
Total	55	76		131

Participant numbers differed across data components. The complete quantitative dataset included all 131 learners, while platform task submissions and responses to the open-ended item were available for smaller groups, as summarised in Table 3.

Table 3 *Participant Numbers by Data Component.*

Data component	n
Learners registered in VLEPIC	131
Learners submitting at least one platform task	125
Complete quantitative dataset: pretest–posttest, IMMS, correlation, and separate and full OLS regression analyses	131
Open-ended responses included in the descriptive qualitative analysis	95

Learning Gains (RQ1)

Learners' performance improved from pretest to posttest after using VLEPIC. As shown in Figure 2, the mean score for the complete sample ($N = 131$) increased from 9.19 in the pretest ($SD = 4.21$, 95% CI [8.46, 9.92]) to 11.92 in the posttest ($SD = 5.70$, 95% CI [10.94, 12.91]), yielding a mean gain of

2.73 points (SD = 2.25, 95% CI [2.34, 3.12]). This upward shift indicates that, at the group level, learners performed better after the intervention.

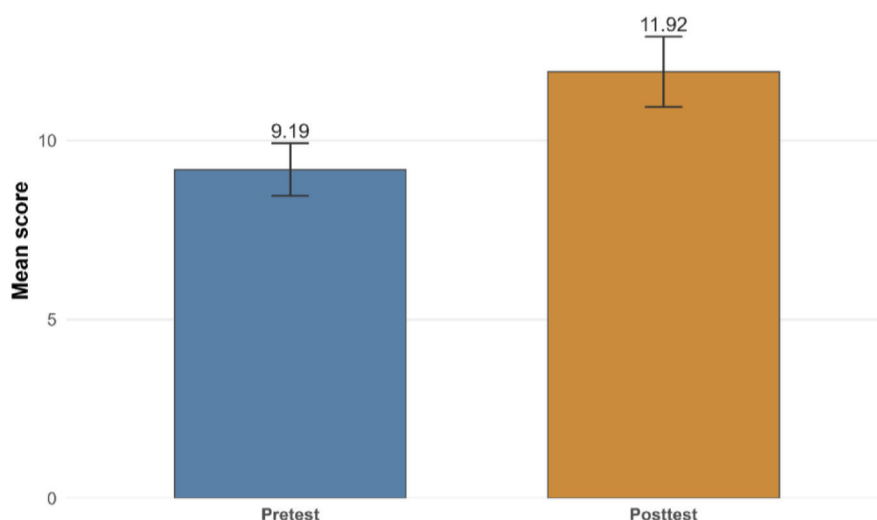


Figure 2 Mean pretest and posttest scores with 95% CIs.

To determine whether this difference was statistically meaningful, gain scores were first examined for normality. Considering they did not meet normality assumptions, a Wilcoxon signed-rank test was used. The test showed a statistically significant difference between pretest and posttest scores, $V = 7336$, $p < .001$. The associated effect size was large ($r = .759$). Of the 131 learners, 110 (84.0%) obtained a higher posttest score, 14 obtained a lower score, and 7 showed no change. The median gain was 3 points (IQR = 2), indicating that the improvement, although modest in absolute score terms, was highly consistent across learners. These results indicate measurable improvement from pretest to posttest after the period of VLEPIC use.

Motivational Profile (RQ2)

Learners reported a positive motivational profile across the four IMMS dimensions, as shown in Table 4 and Figure 3. Mean scores were consistently high and remained close to 4 on a five-point scale, suggesting favourable motivational responses to the learning experience. Among the four dimensions, Relevance ($M = 4.10$, $SD = 0.54$) and Attention ($M = 4.09$, $SD = 0.51$) showed the highest ratings. Confidence ($M = 3.88$, $SD = 0.60$) and Satisfaction ($M = 3.89$, $SD = 0.86$) were slightly lower, although still clearly above the midpoint of the scale.

Table 4 Descriptive Statistics and Ordinal Reliability Estimates for the IMMS ($N = 131$).

IMMS scale	Mean	SD	IQR	95% CI	α	ω
Attention	4.09	0.51	0.63	[4.00, 4.18]	0.81	0.82
Relevance	4.10	0.54	0.67	[4.00, 4.19]	0.78	0.79
Confidence	3.88	0.60	0.78	[3.77, 3.98]	0.80	0.80
Satisfaction	3.89	0.86	1.25	[3.74, 4.04]	0.95	0.96
IMMS total	4.00	0.49	0.67	[3.92, 4.09]	0.93	0.94

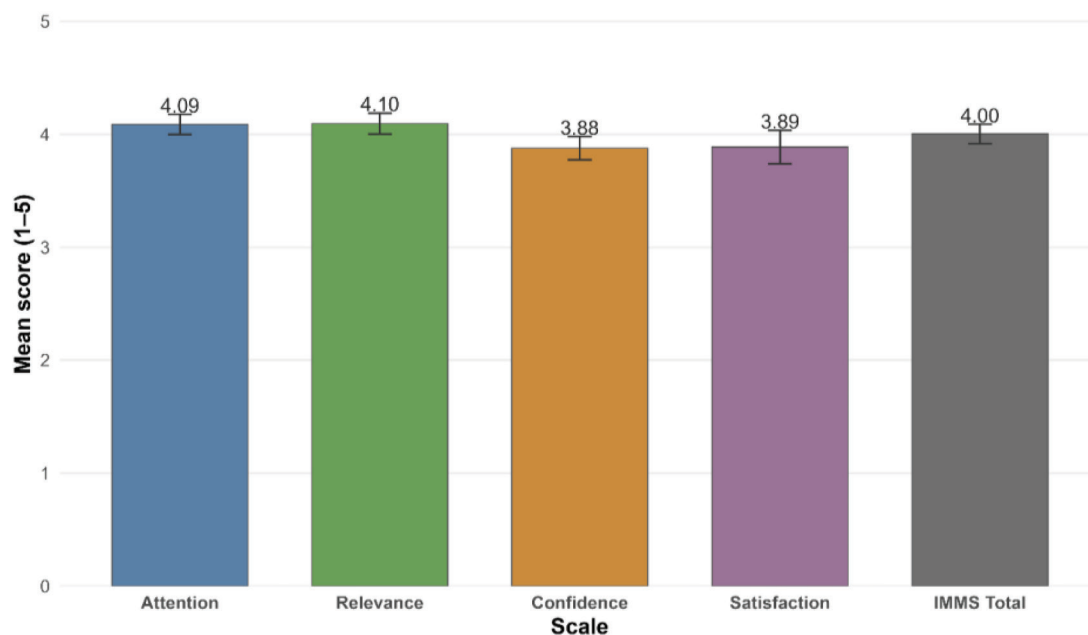


Figure 3 Mean IMMS scores by dimension and total score, with 95% confidence intervals.

The dispersion of scores was relatively moderate for Attention, Relevance, and Confidence, while Satisfaction showed greater variability, as reflected in its higher SD and IQR. CIs for the subscale means remained within the upper range of the scale, reinforcing the overall pattern of positive responses. The total IMMS score was also high ($M = 4.00$, $SD = 0.49$), which supports the view that learners generally perceived the motivational qualities of the environment favourably.

Reliability estimates further supported the consistency of the instrument in this sample. Ordinal alpha and ordinal omega based on polychoric correlations were acceptable to very high across all four subscales and for the total score. Satisfaction showed the strongest internal consistency, followed by the total score, whereas Attention, Relevance, and Confidence also reached acceptable levels. Overall, these results indicate that learners responded positively across the four motivational dimensions and that the IMMS performed consistently in this context.

Additionally, of the 131 participants in the study, 95 provided a response to the open-ended IMMS item. Brief descriptive coding identified recurrent motivating and demotivating categories. Given that some students referred to more than one aspect in the same comment, a single response could be coded into more than one category. Table 5 summarises the most recurrent categories and includes illustrative quotations from students.

Table 5 Recurrent Motivating and Demotivating Aspects Identified in the Open-ended IMMS Item ($n = 95$ Respondents).

Category	Number of mentions	Illustrative quotation
Motivating aspects		
Visual and audiovisual design	34	"I really like the images because they are eye-catching and motivating." (S2)

(Continued)

Category	Number of mentions	Illustrative quotation
Interactive and game-like activities	28	"It helped me that it is interactive, and I learn a lot through the different types of activities." (S3)
Perceived learning and improvement	27	"I liked VLEPIC a lot because I can learn a lot and gain confidence in myself." (S16)
Rewards and mission progression	24	"The points to exchange for rewards motivated me." (S31)
Guidance, clarity, and support	17	"The examples and instructions motivated me." (S8)
Total coded mentions	130	
Demotivating aspects		
Recording audio or video	8	"I don't like recording audios." (S76)
Text-heavy tasks or information load	8	"What demotivates me is the amount of text that sometimes I don't understand." (S54)
Comprehension difficulties	8	"There are some things I don't understand and I need a translator." (S58)
Initial usability difficulties	6	"Its dynamics helped to motivate me, but at the beginning I didn't know how to use it." (S74)
Score-related pressure	4	"I was demotivated when points were taken away after getting a bad mark on a test." (S5)
Total coded mentions	34	

Note. One response could contain more than one coded mention; therefore, the totals exceed the number of students who answered the open-ended question.

Students most often associated motivation with the visual and interactive qualities of VLEPIC, together with a sense of learning, progress, and support. Visual and audiovisual design was the most recurrent motivating aspect, suggesting that images, videos, emojis, and the general appearance of the platform helped to sustain attention and interest. This positive response was reflected in comments such as "I really like the images because they are eye-catching and motivating" (S2) and "The images and videos motivated me" (S60). Interactivity was also repeatedly linked to engagement. One learner explained that "it is interactive, and I learn a lot through the different types of activities" (S3), while another noted that "It felt a lot like a game and was entertaining" (S99). Alongside these features, students frequently referred to perceived learning and improvement. Several comments suggested that VLEPIC was motivating because it was attractive or enjoyable and made learners feel that they were learning more and gaining confidence. As one participant wrote, "I liked VLEPIC a lot because I can learn a lot and gain confidence in myself" (S16), and another stated that "What motivated me most was the possibility of learning more" (S55).

Motivation was also reinforced by progression and support features. Students referred positively to missions, points, rewards, unlocking new stages, and visible advancement through the platform. One learner stated that "Completing the missions motivates me a lot" (S20), while another commented that "While I was advancing in one mission, the next one was unlocked and I was moving up a level" (S40). Guidance also appeared as an important source of support, especially through examples, instructions, and Spanish help. For instance, one student noted that "The examples and instructions motivated me" (S8), and another reported that "What interested me most was the Spanish help and the videos" (S45).

At the same time, some learners identified aspects that reduced their motivation. Recording audios or videos was one of the clearest examples, with comments such as “I don’t like recording audios” (S76) and “I didn’t like recording videos” (S39). Other students mentioned text-heavy tasks, unfamiliar vocabulary, or difficulty understanding parts of the content, as reflected in comments such as “There are some things I don’t understand and I need a translator” (S58) and “There are many words in English and that makes me get stuck” (S96). A smaller number of students referred to initial difficulties in using the platform or to frustration when points were lost after low performance. Overall, these comments complement the positive IMMS profile by showing that motivation in VLEPIC was mainly supported by visual appeal, interactivity, perceived learning, progression, and guidance, whereas recording tasks, textual load, comprehension demands, early navigation difficulties, and score-related pressure reduced motivation for some learners.

Motivation and Learning Gains (RQ3)

The association between learners’ motivational profile and learning gain was examined through Pearson’s correlations between the four IMMS dimensions and the gain score (Table 6). Among the four subscales, only Satisfaction showed a statistically significant positive correlation with learning gain ($r = 0.251$, $p = 0.004$). This bivariate association indicates that higher Satisfaction scores co-occurred with larger learning gains; however, it should not be interpreted as evidence that Satisfaction caused the observed improvement. In contrast, Attention ($r = 0.132$, $p = 0.134$), Relevance ($r = 0.102$, $p = 0.248$), and Confidence ($r = 0.134$, $p = 0.127$) were positively related to gain, but these associations did not reach statistical significance (Figure 4).

Table 6 *Pearson Correlations Between IMMS Dimensions and Learning Gain ($N = 131$).*

IMMS dimension	r	95% CI	p
Attention	0.132	[-0.041, 0.296]	0.134
Relevance	0.102	[-0.071, 0.269]	0.248
Confidence	0.134	[-0.038, 0.299]	0.127
Satisfaction	0.251	[0.083, 0.405]	0.004

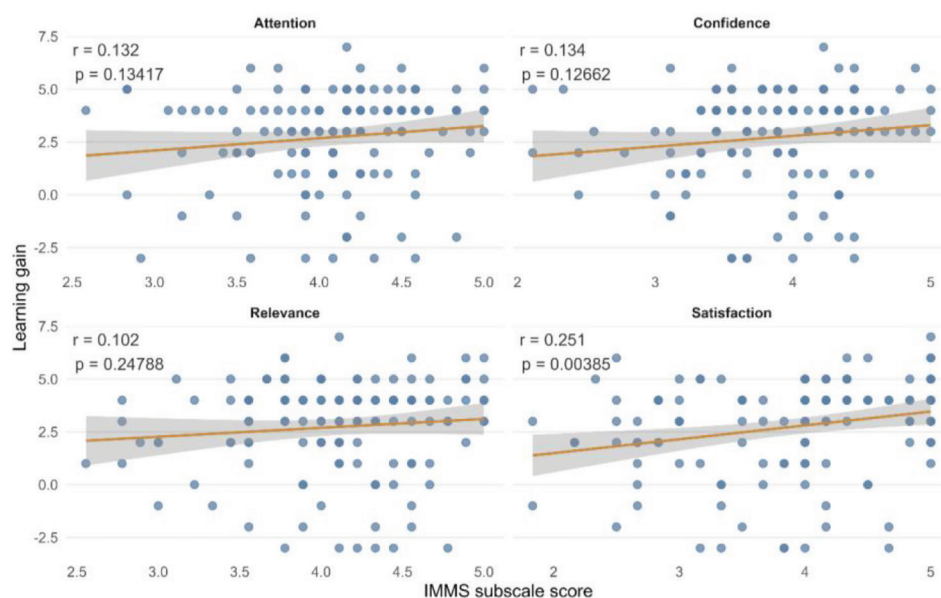


Figure 4 *Associations between IMMS subscales and learning gain.*

To examine whether each motivational dimension was associated with posttest performance after accounting for baseline achievement, four separate OLS regression models were estimated, each including pretest score as a covariate and one IMMS dimension as the motivational predictor (Table 7). The separate adjusted models showed a consistent pattern. All four models were statistically significant overall: Attention, $F(2, 128) = 490.31$, $p < .001$; Relevance, $F(2, 128) = 490.34$, $p < .001$; Confidence, $F(2, 128) = 500.94$, $p < .001$; and Satisfaction, $F(2, 128) = 493.87$, $p < .001$. After controlling for pretest performance, none of the IMMS dimensions was significantly associated with posttest score when entered individually.

Table 7 *Separate OLS Regression Models Predicting Posttest Score from each IMMS Dimension, Controlling for Pretest Score ($N = 131$ for each model).*

IMMS dimension	B	SE	95% CI	t	p	R ²	Adjusted R ²
Attention	0.293	0.336	[-0.372, 0.958]	0.873	.384	.885	.883
Relevance	0.278	0.317	[-0.349, 0.904]	0.877	.382	.885	.883
Confidence	0.512	0.285	[-0.051, 1.075]	1.798	.075	.887	.885
Satisfaction	0.263	0.209	[-0.150, 0.676]	1.260	.210	.885	.883

To examine whether this pattern remained when the four motivational dimensions were considered simultaneously, a full OLS regression model was estimated with posttest score as the dependent variable, the four IMMS dimensions as predictors, and pretest score as a covariate (Table 8). The model was statistically significant, $F(5, 125) = 196.17$, $p < .001$, and explained a large proportion of the variance in posttest performance ($R^2 = .887$, adjusted $R^2 = .882$). Pretest score was the only significant predictor of posttest performance ($B = 1.266$, $SE = 0.044$, $t = 28.799$, $p < .001$), while none of the motivational dimensions reached statistical significance. Multicollinearity diagnostics indicated no concerning levels of collinearity, with VIF values ranging from 1.169 to 2.008 and tolerance values ranging from .498 to .855.

Table 8 *Ordinary Least Squares Regression Model Predicting Posttest Score ($N = 131$).*

Predictor	B	SE	95% CI	t	p
(Intercept)	-1.318	1.576	[-4.438, 1.802]	-0.836	0.405
Pretest	1.266	0.044	[1.179, 1.353]	28.799	< .001
Attention	-0.152	0.473	[-1.088, 0.784]	-0.321	0.748
Relevance	-0.083	0.427	[-0.928, 0.762]	-0.194	0.847
Confidence	0.535	0.398	[-0.253, 1.322]	1.344	0.181
Satisfaction	0.126	0.282	[-0.432, 0.685]	0.448	0.655

Model fit: $R^2 = 0.887$, adjusted $R^2 = 0.882$, $F(5, 125) = 196.17$, $p < .001$.

Discussion

Educational Significance of the Observed Improvement

The findings indicate that VLEPIC was associated with measurable short-term progress in this authentic school context. Learners improved from a pretest mean of 9.19 to a posttest mean of 11.92, with an

average gain of 2.73 points and a large effect size in the pretest–posttest comparison. The large rank-based effect size should be interpreted alongside the modest absolute gain of 2.73 points. In this case, the magnitude of the Wilcoxon effect reflects the strong consistency in the direction of change across learners, with 110 of 131 students improving, and does not imply a correspondingly large increase on the 30-point test scale. Even so, the educational meaning of this result should not be reduced to statistical significance alone. The posttest mean remained below the midpoint of the 30-point scale, so the pattern is better interpreted as evidence of progress, not mastery. In light of this, the main contribution of the study is not to claim strong effectiveness in absolute terms, but to show that the use of a gamified and lightly personalised VLE was compatible with measurable improvement under everyday school conditions.

This interpretation is consistent with recent literature showing that gamification in EFL/ESL and school education is often associated with positive learning and affective outcomes, but that these effects are highly dependent on instructional design, contextual fit, and implementation quality (Chan & Lo, 2024; Dehghanzadeh et al., 2024). Similarly, Simsek and Karakus Yilmaz (2025) examined the relationship between gamification and learning outcomes across subject areas, sample-specific characteristics, implementation forms, and technologies, highlighting the importance of considering implementation context when interpreting the evidence. Chan and Lo's review of empirical EFL/ESL studies concludes that gamification is commonly linked to gains in English proficiency, more positive attitudes, and stronger emotional engagement, while also noting that findings are mixed and context-sensitive. Likewise, Dehghanzadeh et al. (2024) argue that gamification in K–12 settings can support cognitive, affective, and behavioural outcomes, but not in a uniform or automatic way across implementations. From that perspective, the present findings are encouraging, but they do not justify an inflated reading of the gain.

A cautious interpretation is especially important because the study used a one-group pretest–posttest design, and VLEPIC functioned as teacher-assigned reinforcement instead of serving as the sole source of instruction. Under these conditions, it is more defensible to argue that the improvement was consistent with the intervention than to claim that the platform alone caused the learning gains. In our view, that is a more cautious interpretation, because it recognises the educational value of measurable progress in a resource-constrained public-school setting without overstating what the design can demonstrate.

Interpreting the Motivational Profile

The ARCS results suggest that learners experienced VLEPIC positively overall, although not in the same way across all four dimensions. Relevance and Attention were the strongest dimensions, while Confidence and Satisfaction were slightly lower and more dispersed. This pattern suggests that learners responded most positively to the aspects of the VLE related to interest and perceived relevance, while Confidence and Satisfaction were somewhat less consistent across learners. In other words, students were generally drawn into the platform and could recognise its value, yet some still encountered friction when carrying out tasks or judging their own progress. This interpretation is consistent with recent ARCS research, which continues to position Attention, Relevance, Confidence, and Satisfaction as analytically distinct yet interrelated dimensions for explaining how learners respond to instructional design features in technology-enhanced environments (Fang et al., 2024). It is also consistent with current work on gamified LL showing that positive learner responses depend not simply on the presence of game elements, but on how those elements are integrated with task structure, feedback, and pedagogical support (Chan & Lo, 2024; Dehghanzadeh et al., 2024).

The open-ended responses reinforce that interpretation. Learners most often linked their motivation to visual and audiovisual design, interactivity, perceived learning, rewards, mission progression, and guidance, while the most recurrent demotivating aspects were recording tasks, text-heavy activities,

comprehension difficulties, early navigation problems, and pressure related to losing points. This variability is consistent with recent EFL evidence showing that gamified ranking mechanisms can elicit both positive and negative emotional responses across learners (Philpott & Son, 2026). The prominence of rewards and mission progression also aligns with Chan and Lo's (2024) review, which identifies points, badges, rewards, feedback, and progress indicators as recurrent features in gamified EFL/ESL environments. Learners' emphasis on visual design, interactivity, and guidance further supports the view that motivational responses should be interpreted in relation to the overall design of the learning environment (Dehghanzadeh et al., 2024). These qualitative patterns help explain why Attention and Relevance were particularly high, while Confidence and Satisfaction were less stable among learners, and are compatible with recent digital LL research reporting concurrent improvements in language achievement and affective outcomes (Liu et al., 2024).

Motivation and Learning

The relationship between motivation and learning in this study was modest at the bivariate level and became weaker once baseline achievement was included in the adjusted model. At the correlational level, only Satisfaction showed a significant positive association with learning gain, whereas Attention, Relevance, and Confidence showed positive but non-significant associations. The adjusted analyses reinforced this pattern. None of the four IMMS dimensions was significantly associated with posttest performance when entered separately with pretest score as a covariate, and the same result was observed in the full model containing all four dimensions simultaneously. Across these models, baseline achievement remained the strongest predictor of posttest performance.

This pattern is broadly compatible with recent LL research showing that gamification may support learning partly through motivational and affective pathways, while achievement still depends on other learner variables. This differs from findings reported by Shen et al. (2024), where motivation played a mediating role in the relationship between gamification and LL achievement, while Liu et al. (2024) observed concurrent gains in language achievement and affective outcomes. Compared with these studies, the present findings provide a more cautious picture: although learners reported positive motivational responses, these ratings did not explain additional posttest performance once baseline achievement was taken into account. In this study, positive motivational responses seem to have been more clearly reflected in learners' perceptions of the learning experience and perceived progress than in additional explanatory power over posttest achievement.

Conclusion

This study examined the second design iteration of VLEPIC, a gamified and lightly personalised VLE for secondary EFL learners in a public-school context in Ecuador. Regarding RQ1, the findings indicate that the platform was associated with measurable short-term learning progress in this setting. For RQ2, the results suggest a clearly positive motivational experience, especially in terms of attention and relevance, pointing to a learning environment that was perceived as engaging and worthwhile. As for RQ3, the relationship between motivation and learning gain was limited. Satisfaction showed a significant bivariate association with gain, but motivational ratings were not associated with posttest performance once baseline achievement was controlled for.

In general, the study suggests that VLEPIC functioned as a promising reinforcement environment for school-based English learning, particularly because it combined visible progression, gamified structure, and learner support within a resource-constrained context. At the same time, the findings also show that a positive motivational profile does not automatically translate into stronger achievement for all learners, and that differences in starting point still matter.

As part of an ongoing DBR process, this study does not present VLEPIC as a finished solution. Instead, it offers empirical evidence that can inform further refinement of the design and contribute to current discussions on gamification, motivation, and light personalisation in technology-supported English learning. In that sense, the main contribution of the study lies in showing how a school-based VLE can be both pedagogically useful and methodologically informative when examined under real conditions of use.

More specifically, the findings from this iteration point to several priorities for subsequent refinement: clearer re-entry and continuation cues, more visible progress indicators, reduced text density, stronger end-of-sequence support, and more reliable submission workflows. These priorities extend the DBR cycle by translating observed learner difficulties and usage patterns into concrete design decisions for the next version of VLEPIC.

Practical Implications

The findings suggest that gamified VLEs may be especially useful in school-based EFL contexts when they make progress visible and easy to follow. In VLEPIC, learners frequently linked missions, levels, rewards, and unlocked stages to a sense of moving forward, which may be particularly relevant in contexts where access conditions are uneven, and learners need clear indications of their progress.

The qualitative responses also point to areas that matter in practice. While many students valued progression and rewards, others reported difficulties related to recording tasks, textual load, comprehension barriers, early usability issues, and score-related pressure. For teachers and designers, this means that sustaining engagement is not only a matter of adding rewards or progression mechanics. It also requires clear guidance, reduced friction, and tasks that remain manageable as difficulty increases.

The findings also show that light personalisation can be feasibly integrated into a gamified VLE in a resource-constrained school context. In VLEPIC, this personalisation consisted mainly of diagnostic placement and preference-based support within a common learning structure. However, because these features operated alongside gamification, progression, guidance, and other design elements, the present study cannot determine their independent contribution to learning outcomes or motivation. This cautious interpretation is consistent with literature showing that PL encompasses heterogeneous designs and multiple interacting components, while its implementation in secondary education can involve substantial practical complexity (Bernacki et al., 2021; Gunawardena et al., 2024).

Limitations

This study needs to be read in light of several limitations. The evaluation followed a one-group pre-test-posttest design in a single public school in Ecuador, so the findings should be understood within this specific context. The intervention also covered only one school term (3 months) and two quests, which means the study captures a relatively short period of use. In addition, exposure to the platform was not fully uniform across learners. Platform use was episodic and task-oriented, participation varied across the mission sequence, and the available cohort-level logs did not permit exact reconstruction of individual quest completion. The findings therefore reflect VLEPIC under authentic implementation conditions and should not be interpreted as representing equivalent exposure across participants.

There are also limits related to measurement and evidence. Learning was assessed through a researcher-developed pretest and posttest aligned with the content of VLEPIC, so the results reflect improvement within this instructional setting. Motivation was measured through the IMMS, which provided useful information about learners' perceptions, but it was still a post-intervention self-report measure.

The qualitative evidence helped to contextualise the quantitative findings, although it came from a single open-ended survey item and was analysed descriptively. In addition, while the platform combined progression, gamification, guidance, and light personalisation, the study cannot determine the precise contribution of each feature to the observed outcomes.

Future Research

Future research could extend this work by following VLEPIC over a longer period and across additional iterations. This would help to clarify whether the positive motivational profile and short-term learning gains remain stable over time, and whether some effects become stronger, weaker, or more uneven as use continues. It would also be useful to examine more closely how learners with different starting points respond to the platform, especially given the strong role of baseline performance in the present study.

Further work could also look more directly at the contribution of specific design features. In this study, progression mechanics, rewards, guidance, visual design, Spanish support, and preference-based elements operated together, which reflects authentic implementation but makes it difficult to separate their individual influence. Richer qualitative and behavioural evidence could also deepen understanding of how students experience the platform, where they struggle, and how motivation changes during use. Finally, future research could examine teacher orchestration in more depth, since the success of a school-based VLE depends on learner experience and on how the platform is introduced, monitored, and connected to classroom practice.

Use of Generative AI

This manuscript made use of Gemini to assist in grammar correction, readability, and language polishing. No AI tools were used to interpret findings or write the conclusions. The final manuscript was authored and reviewed by the listed authors without additional AI input.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). Standards for educational and psychological testing. *American Educational Research Association*. <https://www.aera.net/publications/books/standards-for-educational-psychological-testing-2014-edition>
- Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189X11428813>
- Arip, A. J., & Hashim, H. (2024). Gamification in English as Second Language (ESL) learning: A systematic review. *International Journal of Academic Research in Progressive Education and Development*, 13(4), 2837–2863. <https://doi.org/10.6007/IJARPED/v13-i4/24069>
- Bernacki, M. L., Greene, M. J., & Lobczowski, N. G. (2021). A systematic review of research on personalized learning: Personalized by whom, to what, how, and for what purpose(s)? *Educational Psychology Review*, 33, 1675–1715. <https://doi.org/10.1007/s10648-021-09615-8>
- Bierer, S. B., Beck Dallaghan, G., Borges, N. J., Brondfield, S., Fung, C. C., Huggett, K. N., Teal, C. R., Thammasitboon, S., & Colbert, C. Y. (2025). Moving beyond simplistic research design in health professions education: What a one-group pretest-posttest design will not prove. *MedEdPORTAL*, 21, 11527. https://doi.org/10.15766/mep_2374-8265.11527
- British Educational Research Association. (2024). *Ethical guidelines for educational research* (5th ed.). <https://www.bera.ac.uk/publication/ethical-guidelines-for-educational-research-fifth-edition-2024>

- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Houghton Mifflin. <https://www.sfu.ca/~palys/Campbell&Stanley-1959-Exptl&QuasiExptlDesignsForResearch.pdf>
- Chan, S., & Lo, N. (2024). Enhancing EFL/ESL instruction through gamification: A comprehensive review of empirical evidence. *Frontiers in Education*, 9, 1395155. <https://doi.org/10.3389/educ.2024.1395155>
- Creswell, J. W. (2022). *A concise introduction to mixed methods research* (2nd ed.). SAGE Publications.
- Creswell, J. W., & Plano Clark, V. L. (2017). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publications.
- Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaei, E., & Noroozi, O. (2021). Using gamification to support learning English as a second language: A systematic review. *Computer Assisted Language Learning*, 34(7), 934–957. <https://doi.org/10.1080/09588221.2019.1648298>
- Dehghanzadeh, H., Farrokhnia, M., Dehghanzadeh, H., Taghipour, K., & Noroozi, O. (2024). Using gamification to support learning in K-12 education: A systematic literature review. *British Journal of Educational Technology*, 55(1), 34–70. <https://doi.org/10.1111/bjet.13335>
- Design-Based Research Collective. (2003). Design-based research: An emerging paradigm for educational inquiry. *Educational Researcher*, 32(1), 5–8. <https://doi.org/10.3102/0013189X032001005>
- Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011). From game design elements to gamefulness: Defining “gamification”. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). Association for Computing Machinery. <https://doi.org/10.1145/2181037.2181040>
- Di Serio, A., Ibáñez, M. B., & Delgado Kloos, C. (2013). Impact of an augmented reality system on students’ motivation for a visual art course. *Computers & Education*, 68, 586–596. <https://doi.org/10.1016/j.compedu.2012.03.002>
- du Plooy, E., Casteleijn, D., & Franzsen, D. (2024). Personalized adaptive learning in higher education: A scoping review of key characteristics and impact on academic performance and engagement. *Heliyon*, 10(21), e39630. <https://doi.org/10.1016/j.heliyon.2024.e39630>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399–412. <https://doi.org/10.1111/bjop.12046>
- Fang, X., Ng, D. T. K., Leung, J. K. L., & Xu, H. (2024). The applications of the ARCS model in instructional design, theoretical framework, and measurement tool: A systematic review of empirical studies. *Interactive Learning Environments*, 32(10), 5919–5946. <https://doi.org/10.1080/10494820.2023.2240867>
- Fiel Peres, F. (2026). Effect sizes for nonparametric tests. *Biochemia Medica*, 36(1), 010101. <https://doi.org/10.11613/BM.2026.010101>
- Forsström, S., Njå, M., Munthe, E., Álvarez-Galván, J.-L., & Houldsworth, L. (2025). *The impact of digital technologies on students’ learning: Results from a literature review* (OECD Education Working Papers No. 335). OECD Publishing. <https://doi.org/10.1787/9997e7b3-en>
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation*, 17(3), 1–13. <https://doi.org/10.7275/n560-j767>
- Gunawardena, M., Bishop, P., & Aviruppola, K. (2024). Personalized learning: The simple, the complicated, the complex and the chaotic. *Teaching and Teacher Education*, 139, 104429. <https://doi.org/10.1016/j.tate.2023.104429>
- Hsieh, H.-F., & Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qualitative Health Research*, 15(9), 1277–1288. <https://doi.org/10.1177/1049732305276687>
- Kaur, P., Kumar, H., & Kaushal, S. (2023). Technology-assisted language learning adaptive systems: A comprehensive review. *International Journal of Cognitive Computing in Engineering*, 4, 301–313. <https://doi.org/10.1016/j.ijcce.2023.09.002>

- Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. Springer. <https://doi.org/10.1007/978-1-4419-1250-3>
- Lafleur, L. (2024). The effects of gamified daily awards on digital vocabulary flashcard learning: A case study. *Technology in Language Teaching & Learning*, 6(2), 1195. <https://doi.org/10.29140/tltl.v6n2.1195>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4, 863. <https://doi.org/10.3389/fpsyg.2013.00863>
- Liu, G.-Z., Fathi, J., & Rahimi, M. (2024). Using digital gamification to improve language achievement, foreign language enjoyment, and ideal L2 self: A case of English as a foreign language learners. *Journal of Computer Assisted Learning*, 40(4), 1347–1364. <https://doi.org/10.1111/jcal.12954>
- Loorbach, N., Peters, O., Karreman, J., & Steehouder, M. (2015). Validation of the Instructional Materials Motivation Survey (IMMS) in a self-directed instructional setting aimed at working with technology. *British Journal of Educational Technology*, 46(1), 204–218. <https://doi.org/10.1111/bjet.12138>
- Lu, J., Schmidt, M. M., & Shin, J. (2025). *Beyond technological usability: Exploratory factor analysis of the comprehensive assessment of usability scale for learning technologies (CAUSLT)* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2501.18754>
- McKenney, S., & Reeves, T. C. (2018). *Conducting educational design research* (2nd ed.). Routledge. <https://doi.org/10.4324/9781315105642>
- Ministerio de Educación del Ecuador. (2016). *Currículo de los niveles de educación obligatoria: Subnivel Superior (English as a Foreign Language)*. <https://educacion.gob.ec/curriculo-lengua-extranjera/>
- Philpott, A., & Son, J.-B. (2026). EFL students' emotional reactions to class leaderboards. *Technology in Language Teaching & Learning*, 8, 102623. <https://doi.org/10.29140/tltl.2026.102623>
- Ramírez-Montoya, M.-S., & Lugo-Ocando, J. (2020). Systematic review of mixed methods in the framework of educational innovation. *Comunicar*, 28(65), 9–20. <https://doi.org/10.3916/C65-2020-01>
- Sandelowski, M. (2000). Whatever happened to qualitative description? *Research in Nursing & Health*, 23(4), 334–340. [https://doi.org/10.1002/1098-240X\(200008\)23:4<334::AID-NUR9>3.0.CO;2-G](https://doi.org/10.1002/1098-240X(200008)23:4<334::AID-NUR9>3.0.CO;2-G)
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shang, X. (2025). Mapping the integration of gamification into foreign language learning in E-learning settings: A bibliometric analysis. *Results in Engineering*, 28, 108452. <https://doi.org/10.1016/j.rineng.2025.108452>
- Shemshack, A., & Spector, J. M. (2020). A systematic literature review of personalized learning terms. *Smart Learning Environments*, 7, 33. <https://doi.org/10.1186/s40561-020-00140-9>
- Shen, Z., Lai, M., & Wang, F. (2024). Investigating the influence of gamification on motivation and learning outcomes in online language learning. *Frontiers in Psychology*, 15, 1295709. <https://doi.org/10.3389/fpsyg.2024.1295709>
- Simsek, E., & Karakus Yilmaz, T. (2025). A systematic review of the effects of gamification in online learning environments on learning outcomes. *Open Praxis*, 17(1), 166–183. <https://doi.org/10.55982/openpraxis.17.1.692>
- Sireci, S. G., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100–107. <https://doi.org/10.7334/psicothema2013.256>
- Tashakkori, A., & Teddlie, C. (Eds.). (2010). *SAGE handbook of mixed methods in social & behavioral research* (2nd ed.). SAGE Publications. <https://doi.org/10.4135/9781506335193>
- Velarde Orozco, M. T., & de Benito Crosetti, B. L. (2025). Enhancing English language teaching and user experience in virtual environments: A systematic review on gamification and personalised

- learning. *Journal of Language and Education*, 11(2), 157–174. <https://doi.org/10.17323/jle.2025.24798>
- Velarde Orozco, M. T., & de Benito Crosetti, B. L. (2026a). Evaluating a virtual learning environment for secondary English in a public school: Usability, motivation, and engagement. *Education Sciences*, 16(1), 169. <https://doi.org/10.3390/educsci16010169>
- Velarde Orozco, M. T., & de Benito Crosetti, B. L. (2026b). VLEPIC: Interaction design for secondary English in a gamified and personalised virtual learning environment. *Digital*, 6(3), 56. <https://doi.org/10.3390/digital6030056>
- Vickers, A. J., & Altman, D. G. (2001). Statistics notes: Analysing controlled trials with baseline and follow up measurements. *BMJ*, 323(7321), 1123–1124. <https://doi.org/10.1136/bmj.323.7321.1123>
- World Medical Association. (2024). *WMA Declaration of Helsinki: Ethical principles for medical research involving human participants*. <https://www.wma.net/policies-post/wma-declaration-of-helsinki/>
- Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, 13, 1030790. <https://doi.org/10.3389/fpsyg.2022.1030790>
- Züll, C. (2016). *Open-ended questions* (GESIS Survey Guidelines). GESIS – Leibniz Institute for the Social Sciences. https://doi.org/10.15465/gesis-sg_en_002