https://doi.org/10.17323/jle.2025.24798

Enhancing English Language Teaching and User Experience in Virtual Environments: A Systematic Review on Gamification and Personalised Learning

Myriam Tatiana Velarde Orozco, Bárbara Luisa De Benito Crosetti

University ot the Balearic Islands, Spain

ABSTRACT

Introduction: Virtual learning environments (VLEs) have become central to English language teaching (ELT), although persistent disengagement suggests that design must go beyond content delivery. Gamification and personalised learning (PL) contribute to enhanced user experience (UX) and better learning outcomes, but evidence on their combined effects remains fragmented.

Purpose: This systematic literature review (SLR) explains how gamification and PL influence UX, motivation, engagement and achievement in ELT-oriented VLEs, identifies effective design practices, and maps the implementation challenges that constrain them.

Method: Following PRISMA-2020 guidelines, 46 empirical studies (2015-2024) were retrieved from Scopus, Web of Science, ERIC, and Dialnet. Two extraction matrices captured bibliographic, contextual and analytical data; methodological quality was appraised with MMAT-2018. Comparative and narrative syntheses linked design features to four outcome clusters: motivation/engagement, UX, academic performance, and learner satisfaction.

Results: Challenge-based game mechanics, points and rewards reliably increased motivation and engagement, especially when integrated with adaptive feedback. PL strategies (adaptive difficulty, learner-directed paths and tailored content) produced the strongest dual gains in satisfaction and achievement. High UX emerged only when interfaces minimised cognitive load and feedback was timely. Gaps persist between short-term motivational spikes and durable learning, and competitive elements induce anxiety in novices. Key obstacles include limited digital literacy, bandwidth constraints and sparse reporting on implementation fidelity.

Conclusion: Gamification and PL can substantially enhance UX and selected learning outcomes in ELT-oriented VLEs, but only when designs align with curricular goals, resource realities and learner profiles. Future research should pursue longer mixed-method trials, transparent adaptivity and scalable models for low-resource contexts.

KEYWORDS

educational technology, English language teaching, gamification, personalised learning, user experience, virtual learning environments

Citation: 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

Correspondence:
Myriam Tatiana Velarde Orozco,
myriam-tatiana.velarde@uib.cat

Received: February 28, 2025
Accepted: June 10, 2025
Published: June 30, 2025

INTRODUCTION

Virtual learning environments (VLEs) have progressed from peripheral digital tools to essential platforms for language education across all levels of formal schooling. By integrating extensive resource repositories, synchronous and asynchronous communication channels, and automated assessment functions, VLEs provide flexibility and reach unavailable in conventional classrooms (Al-Busaidi & Al-Shihi, 2012). However, the rapid expansion of online instruction has exposed serious shortcomings. A large international survey conducted during the COVID-19 pandemic revealed that 72% of learners felt only "slightly" or "not at all" engaged with their online courses, linking this disengagement to weaker academic outcomes (Hollister et al., 2022). This evidence raises a critical question:

how can VLEs be designed to go beyond content delivery and genuinely engage learners?

User experience (UX) has emerged as a decisive factor in meeting this challenge. Intuitive navigation, logically organised information, and timely, responsive feedback reduce cognitive load, allowing learners to focus on the instructional task; these features consistently predict lower attrition and better performance in digital settings (Cho et al., 2022; Sanchis-Font et al., 2021). Moreover, two design approaches that could deserve particular attention within the broader UX agenda are gamification and personalised learning (PL).

Gamification introduces game elements (points, competition, narrative, etc.) into educational contexts to spark and sustain motivation (Deterding et al., 2011). Systematic reviews in English language teaching (ELT) report increased situational interest and modest improvements in vocabulary and communicative competence when game mechanics are carefully aligned with learning goals (Hung et al., 2018; Zainuddin et al., 2020; Zhang & Hasim, 2023). These reviews also note that overused or poorly integrated game features can lose their appeal, particularly when they are not pedagogically sound.

PL, in contrast, adapts content, pacing, and feedback to the needs of each learner. It is broader than adaptive learning (AL), which relies on algorithms to adjust difficulty or sequencing in real time (Klašnja-Milićević et al., 2011; Shemshack & Spector, 2020). In this review, we adopt PL as an umbrella term and include studies labelled "adaptive" when they demonstrably provide individualised support (Bernacki et al., 2021).

Although gamification and PL can address aspects of the UX challenge, their research traditions have largely evolved in parallel. Gamification reviews emphasise motivational and linguistic outcomes but rarely tackle personalisation or nuanced UX metrics (Hung et al., 2018; Zainuddin et al., 2020; Zhang & Hasim, 2023). Conversely, PL syntheses focus on analytical adaptation, often in STEM or general-education contexts, and seldom consider game elements, with only occasional reference to language learning (LL) applications (Ali et al., 2024; du Plooy et al., 2024; Gevorgyan, 2024). To our knowledge, no review has systematically examined how the combined use of gamification and PL shapes UX and learning outcomes in ELT-oriented VLEs.

Responding to this underexplored area, the present systematic literature review (SLR) examines empirical investigations published between 2015 and 2024 on the joint implementation of gamification and PL in VLEs designed for ELT. It seeks to clarify design features, evaluate their impact on learners, and identify methodological patterns and challenges. The review is guided by four research questions (RQs):

RQ1: What design features and pedagogical strategies are most effective in enhancing UX in ELT-oriented VLEs?

RQ2: How does gamification influence student motivation/engagement and academic performance in these environments?

RQ3: Which PL strategies are commonly employed, and how do they affect educational outcomes and learner satisfaction?

RQ4: What implementation challenges are reported, particularly in low-resource settings or among users with limited digital competence?

By answering these questions, we aim to guide educators, instructional designers, and researchers in designing VLEs that are not merely functional but pedagogically responsive and engaging for diverse language learners.

LITERATURE REVIEW

This section integrates three strands: UX in educational technology (ET), gamification, and PL strategies in ELT, to frame the RQs and establish the theoretical foundations needed to understand how design choices in VLEs influence motivation, engagement, and learning outcomes in ELT.

Conceptual Foundations of UX in ET

UX in ET can be framed by three complementary theoretical lenses. First, the Unified Theory of Acceptance and Use of Technology 2 (UTAUT2) posits that learners' intentions to continue using an online system depend on performance and effort expectancy as well as hedonic motivation, among other factors (Venkatesh et al., 2012). These constructs correspond directly to interface qualities such as clear navigation paths and responsive feedback that reduce friction in LL platforms. Second, Self-Determination Theory emphasises autonomy, competence and relatedness as drivers of intrinsic motivation (Ryan & Deci, 2000). Practically, VLE features that grant students meaningful choices (e.g. selecting task sequences) or provide competence-affirming feedback can satisfy these needs and sustain engagement during repetitive language practice. Third, engagement frameworks rooted in educational psychology describe behavioural, cognitive and emotional components that must be balanced for deep learning (Fredricks et al., 2004). Cognitive-load research demonstrates that poorly organised information architectures induce extraneous load and suppress these engagement dimensions (Sweller, 2011).

Recent analyses of learner forum data confirm that user control and timely system responses are strong affective

predictors of positive UX (Sanchis-Font et al., 2020). These models clarify why intuitive interfaces, motivational design and load-reducing layouts are not solely aesthetic additions but theoretical prerequisites for effective virtual LL.

Gamification in VLEs for ELT

Gamification, defined as "the application of game-design elements in non-game contexts" (Deterding et al., 2011, p. 10), is usually organised around achievement, progression, and social-interaction mechanics (Majuri et al., 2018). A meta-synthesis of 128 studies identified medium-sized advantages in learner motivation and accomplishment but noted that only a small minority (14%) used validated UX instruments, limiting insight into experience quality (Koivisto & Hamari, 2019). The same pattern appears in ELT reviews. Helvich et al. (2023) and Zhang and Hasim (2023) found that most trials lasted fewer than five weeks and relied on ad hoc engagement measures. Nonetheless, controlled classroom evidence is emerging; an eight-week study with Thai EFL undergraduates that combined a branching narrative with points and levels led participants to outperform a non-gamified comparison group on oral-fluency and vocabulary post-tests (Wichadee & Pattanapichet, 2018). Two consistent cautions arise; motivational novelty fades when mechanics are repetitive, and misaligned game tasks can trigger cognitive overload (Helvich et al., 2023; Koivisto & Hamari, 2019). Effective gamification in ELT-oriented VLEs therefore requires varied, curriculum-integrated mechanics and systematic UX monitoring.

PL and AL in VLEs for ELT

PL in virtual ELT platforms is driven by an explicit learner profile, interests, prior knowledge, and self-set goals, against which teachers or students manually select tasks, resources and pacing; common tools include choice boards, goal contracts and self-assessment rubrics that shape weekly paths (Bernacki et al., 2021; Pane, 2017; Walkington & Bernacki, 2020). AL embeds algorithms that adjust difficulty on the fly: spaced-repetition decks resurface forgotten words, speech-recognition engines reintroduce problematic phonemes and rule-based tutors reorder grammar drills when error rates spike (Chaichumpa et al., 2021; Klašnja-Milićević et al., 2011; Nazempour & Darabi, 2023).

Meta-analyses indicate that PL mainly boosts motivation through perceived agency, whereas AL induces small-to-moderate advances in vocabulary and pronunciation when adaptation rules are transparent (du Plooy et al., 2024; Gevorgyan, 2024). Li et al. (2022) found that giving students freedom to choose their own projects, while an algorithm quietly fine-tuned tasks in real time, helped them master the target language forms faster than either strategy on its own. This interplay of meaningful choice and instant adjustment appears to foster a richer UX. Consistent with the conceptual boundaries outlined in the introduction, AL studies are analysed here within the broader PL corpus considering that both emphasise individualised instructional support.

Previous Studies and Gaps in the Literature

Despite the rapid growth of research on digital LL, three persistent gaps remain: conceptual fragmentation, methodological limitations, and contextual blind spots. Conceptual fragmentation is critical; gamification studies focus on motivational benefits (Koivisto & Hamari, 2019; Zhang & Hasim, 2023), while PL/AL reviews emphasise algorithmic tailoring (Ali et al., 2024; Gevorgyan, 2024). Virtually no primary study or synthesis examines how the two logics interact to improve UX.

Methodological weaknesses also compound the problem as interventions are typically brief (five to eight weeks), rely on self-developed or unvalidated instruments, and focus narrowly on receptive vocabulary, leaving productive skills and rigorously validated outcome measures underexplored (Helvich et al., 2023). Ali et al. (2024) further highlight heterogeneous methods, scarce triangulation of mixed data, and sparse documentation of implementation fidelity across PL trials. Contextual blind spots also persist, since evidence remains limited for bandwidth-constrained VLEs or learners with low digital literacy, even though studies consistently flag poor connectivity and limited technical skills as key barriers (Helvich et al., 2023).

This SLR responds directly to these limitations. RQ1 charts effective design features by jointly analysing gamification and PL. RQ2 and RQ3 analyse their separate and combined impacts on motivation/engagement and academic achievement. While UX was not always the explicit focus, studies were included if they explored gamification or PL and provided at least one relevant indicator (whether self-reported, behavioural or performance-based), even if not derived from validated instruments. RQ4 compares implementation challenges, with attention to low-resource contexts and novice users. In doing so, the review offers what may be the first integrated, methodologically focused map of how gamified-personalised VLEs shape the full UX continuum in ELT.

METHOD

This study employed the methodology of a SLR. Marín-Juarros (2022) claims that an SLR refers to an exhaustive examination of the literature using systematic methods that allow for replication and updating, addressing one or more RQs by means of a secondary study that combines the results of primary studies. An SLR reveals gaps, deficiencies, and trends in the existing evidence, providing a foundation for and guiding future research in the field (Munn et al., 2018).

Protocol

This SLR was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). An a priori protocol was developed to guide the conduct of this review, outlining the objectives, eligibility criteria, search strategy, study selection process, data extraction methods, quality assessment approach, and methods for data synthesis.

Eligibility Criteria

Studies were eligible for inclusion in this review when all the following conditions were met:

1. Empirical studies that examine gamification and/or PL implemented primarily in ELT-oriented VLEs and that report at least one of the following outcomes: UX, student motivation and/or engagement, learner satisfaction, or academic performance.
2. Articles published between January 2015 and July 2024 in English or Spanish.
3. Peer-reviewed journal articles. Grey literature (conference papers, proceedings, dissertations, theses, book chapters, reports) was excluded.
4. Any empirical design (qualitative, quantitative, or mixed methods). Reviews, theoretical papers, opinion essays, and editorials were excluded.
5. Full-text accessible through open access or the University of the Balearic Islands library.

Studies not meeting all the above criteria were excluded.

Information Sources

The literature search was conducted in the following electronic databases: Scopus, ISI Web of Science (WoS), ERIC, and Dialnet. The final database search was completed on 17th July 2024.

Search Strategy

A comprehensive search strategy was developed and employed to identify all relevant studies. It included a set of predefined keywords and synonyms related to UX, gamification, PL and ELT. These terms were combined using Boolean operators to set out the search strings (Table 1) and maximise the retrieval of relevant studies.

Study Selection Process

Initial database searches returned 4673 records. Filters specific to each database, publication year (2015 onwards), language (English or Spanish), and publication type (journal articles) were applied to eliminate results that did not meet the inclusion criteria, together with subject-area filters to retain studies relevant to the field. After this first screening, 90 articles were selected for further review and imported into RefWorks for organisation and deduplication. After removing two duplicates, 88 unique records were imported into Parsif.al to facilitate the screening and eligibility process. The selection process comprised two main stages: a title-and-abstract screening and a full-text assessment.

The titles and abstracts of the 88 unique articles were screened against the predefined eligibility criteria. Articles that clearly did not meet them were excluded at this stage. After the title-and-abstract screening, the full texts of the remaining records were retrieved and examined against eligibility criteria. 39 papers were excluded at this stage; some because the full text was unavailable, others since a closer reading showed they did not address the review questions, and finally 13 as they were theoretical, review, or editorial papers rather than empirical studies.

Table 1

Search strings for database query

In Spanish	In English
(“experiencia de usuario” OR UX OR “experiencia del estudiante” OR gamificación OR “juegos educativos” OR “juegos serios” OR “personalización del aprendizaje” OR “aprendizaje personalizado”) AND (“entornos virtuales de aprendizaje” OR “e-learning” OR “plataformas educativas virtuales”) AND (“enseñanza del inglés” OR “aprendizaje del inglés” OR EFL OR ESL) AND (“motivación de los estudiantes” OR engagement OR “rendimiento académico” OR “logro académico” OR “satisfacción de los estudiantes”)	(“user experience” OR UX OR “student experience” OR gamification OR “educational games” OR “serious games” OR “personalised learning” OR “learning personalisation”) AND (“virtual learning environments” OR e-learning OR “online learning platforms”) AND (“English teaching” OR “English learning” OR EFL OR ESL) AND (“student motivation” OR engagement OR “academic performance” OR “academic achievement” OR “student satisfaction”)
For Dialnet (160-character limit)	
(“experiencia de usuario” OR UX OR gamificación OR “personalización del aprendizaje”) AND (“entornos virtuales de aprendizaje” OR e-learning) AND (“enseñanza del inglés” OR EFL) AND (motivación OR “rendimiento académico”)	(“user experience” OR UX OR “student experience” OR gamification OR “personalized learning”) AND (“virtual learning environments” OR e-learning) AND (“English teaching” OR ESL) AND (“student motivation” OR engagement OR “academic performance”)

The full-text appraisal therefore resulted in a final sample of 46 empirical investigations. The numbers screened, assessed and excluded at each step, together with the reasons for exclusion, are illustrated in the PRISMA 2020 flow diagram (Figure 1). After the title-and-abstract screening stage, to support internal organisation and data management throughout the review, each article was assigned a unique reference code based on the year of publication (e.g., A for 2015, B for 2016, etc.), and the number of order (e.g., A1, A2, B1, etc.). As 13 studies were excluded during the full-text appraisal, some codes in the final sample appear non-consecutive.

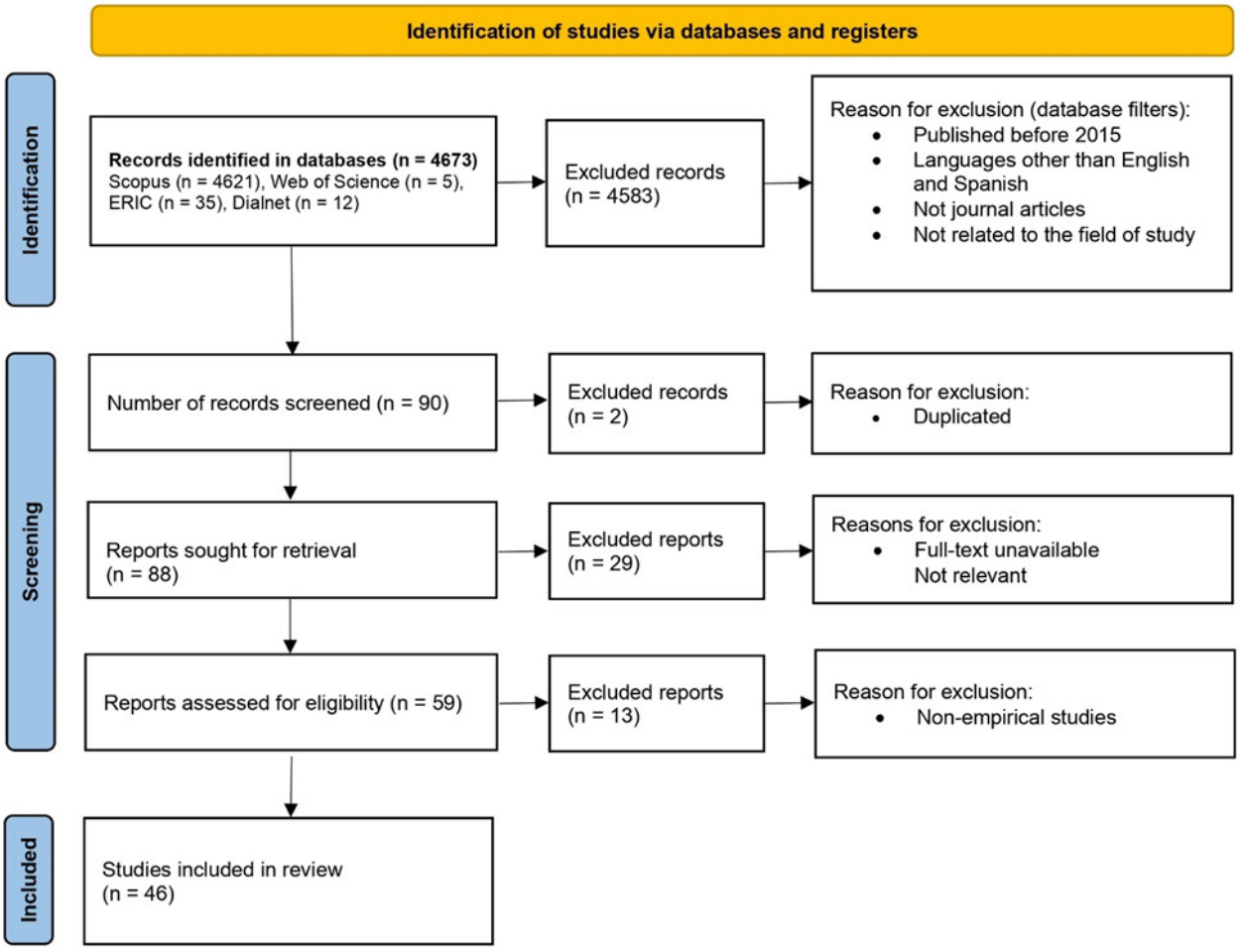
Data Extraction

A structured data-extraction strategy was applied to each of the 46 eligible studies. Two complementary matrices, created in Microsoft Excel, were used to capture both descriptive and analytical information. The descriptive matrix included: Reference code, Title, Authors, Year, Journal, DOI or link, Database, Country/educational level/context, Sample size, Participant profile, Study design, and Study type. The analytical matrix recorded: Reference code, VLE description/platform, Gamification features and pedagogy, Personalisation/adaptive strategies and pedagogy, Duration/intensity of intervention, UX measurement and findings, Motivation, Engagement, Academic performance, Learner satisfaction, Reported links between design strategies and outcomes, Reported implementation challenges, Main conclusions related to the RQs, and quality verdict.

Both matrices were piloted on five studies and adjusted for clarity before full extraction began. All data were extracted from the full-text reports and each field was verified against the original articles. Any discrepancies were resolved through discussion and, where necessary, re-consultation of the source study. No automation tools were employed during data collection. The full descriptive and analytical

Figure 1

PRISMA Flowchart



matrices are presented in Appendix A and Appendix C, respectively.

Quality Assessment and Risk of Bias

The methodological quality and risk of bias for each of the 46 included studies were assessed using the Mixed Methods Appraisal Tool (MMAT), 2018 version (Hong et al., 2018). The MMAT was selected as a validated instrument suitable for appraising diverse study designs (qualitative, quantitative, and mixed methods) present in this review.

Following the MMAT protocol, after confirming suitability via two screening questions, the five relevant design-specific criteria for each study were rated as Yes (met), No (not met), or Can't tell (insufficient information). An overall methodological quality verdict (High, Moderate or Low) was then qualitatively derived for each study with a detailed rationale. This judgement was based on a critical consideration of the pattern of responses, particularly for criteria essential to each study design, rather than a numerical score or arbitrary threshold, aligning with MMAT developer recommendations. These quality assessments were used to understand the general strength of the evidence base, inform the data synthesis, and discuss limitations. The full MMAT assessment for each study is provided in Appendix B.

Data Synthesis

The synthesis was based on the descriptive and analytical data previously extracted, focusing on the relationship between design logics and educational outcomes. This mapping provided a general profile of the evidence base and guided the construction of a comparative matrix that crossed the two focal design logics, gamification mechanics and personalisation strategies, with four core outcome clusters: motivation/engagement, UX, academic performance, and learner satisfaction. Motivation and engagement were analysed jointly due to their frequent conceptual and operational overlap in the included studies. To support the analysis, a heatmap visualisation (generated in R) was produced. Descriptive details (educational level, participant profile, platform, intervention length) were also noted to frame the results, though not analysed separately.

Implementation challenges were collected across studies and sorted into three practical categories: technical, pedagogical, and learner-related, considering local resource levels and digital competence whenever possible. Throughout the analysis, MMAT quality ratings served as a reference point, with lower-quality studies flagged but not removed. In combination, the matrix, heatmap visualisation, and challenge summary underpin the structure of the Results section.

RESULTS

Study Characteristics

46 empirical studies met the final eligibility criteria (2015–2024; median publication year = 2022). Most were published in or after 2021 (74%), indicating a marked recent increase in interest in gamified-personalised ELT-oriented VLEs. Most research was carried out in Asia (26), followed by Europe (12), North/South America (7), and Africa (1). The university sector dominated the corpus (32), with a smaller representation from primary/secondary settings (10) and early-childhood or teacher-training contexts (4).

23 studies adopted purely quantitative designs but only two were randomised controlled trials (RCTs), 18 used mixed methods, and 5 were qualitative. MMAT appraisal classified 20 studies as high quality, 24 as moderate, and 2 as low. Where available, sample sizes ranged from small classroom cohorts to institution-wide groups, most commonly between 50 and 200 learners. A detailed tabulation of the characteristics of each included study is presented in Appendix A.

Core Synthesis: Comparative Matrix and Co-Occurrence Pattern

A single comparative matrix cross-links all gamification mechanics and personalisation strategies with four outcome clusters (motivation/engagement, UX, academic performance, and learner satisfaction). This comprehensive table (Appendix D) reports the number of studies, the predominant effect (+, +/–, –), and one high-quality exemplar study, where available; note that some design logics were only identified in few studies that did not meet high-quality criteria. For a more accessible overview of the most salient findings, a summarised version is presented in Table 2.

To complement the comparative matrix (Appendix D) and facilitate the visual identification of co-occurrence patterns and the strength of evidence, the data are also represented in a heatmap bubble chart (Figure 2). In this visualisation, the size of each bubble is proportional to the number of studies that investigated the specific combination of a design logic and an outcome cluster, whilst the colour of the bubble indicates the predominant direction of effect, as defined in the accompanying legend.

Design Features and Pedagogical Strategies Enhancing UX (RQ1)

The analysis confirms that no single design logic guarantees a universally positive UX in ELT-oriented VLEs. A small set of design logics emerged as consistently beneficial for UX.

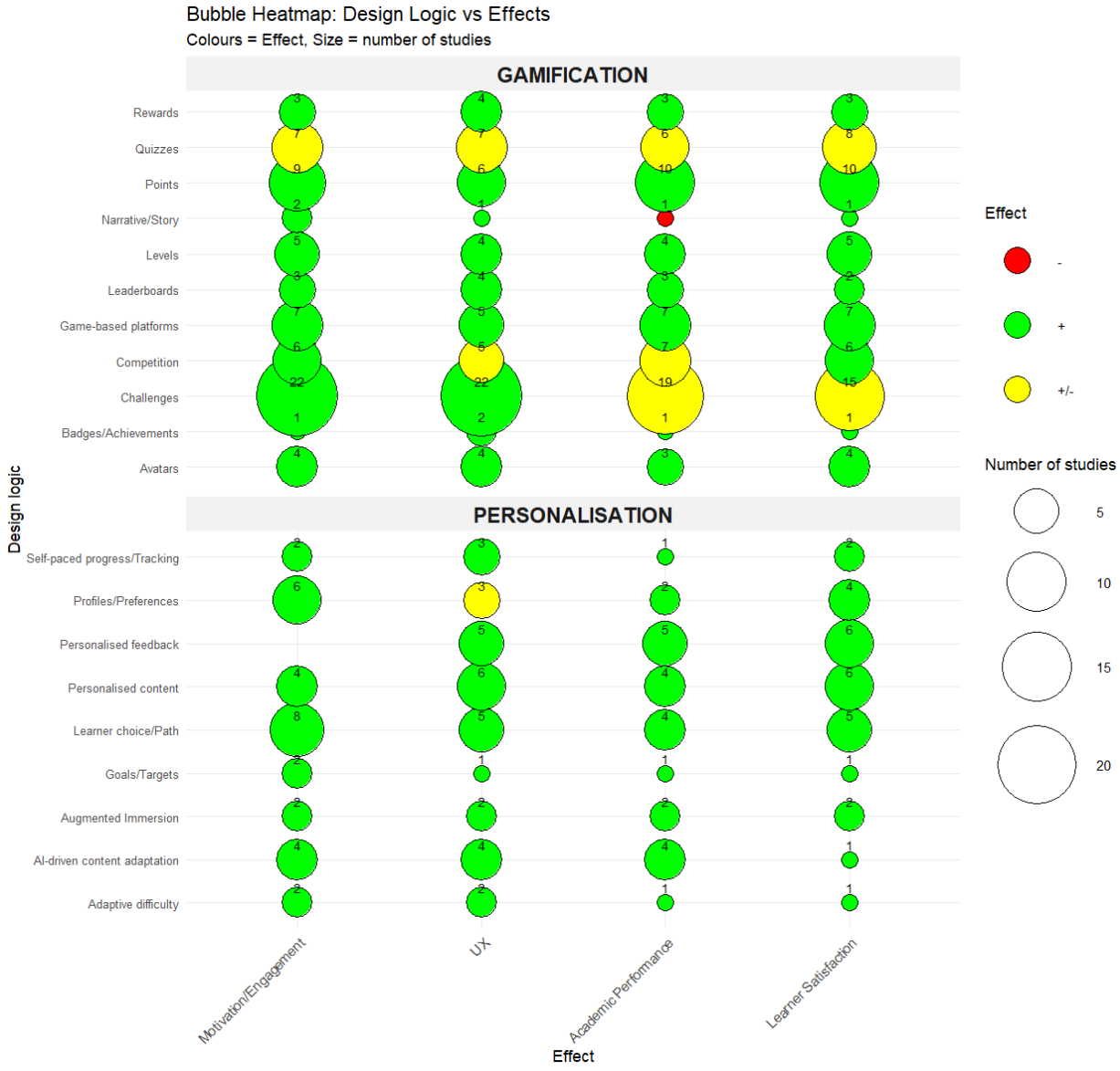
Table 2

Summary of Key Design Logics and their Impact on ELT Outcomes in VLEs.

Design Category	Key Design Logics	Impact on Motivation/ Engagement & UX	Impact on Academic Performance & Learner Satisfaction	Pedagogical Implications
High-impact gamification	Challenges, game-based platforms, points, rewards, levels, avatars, quizzes	Consistently positive	Generally positive, though challenges show mixed results	Effective for engagement; best when integrated with scaffolding
Mixed-result gamification	Competition, leaderboards	Context-dependent; can be demotivating	Mixed results	Requires thoughtful design to prevent exclusion
Consistent personalisation	Learner choice/path, personalised content, personalised feedback	Consistently positive	Consistently positive	Essential for adapting to learner needs
Emerging UX technologies	Artificial intelligence (AI) driven adaptation, augmented immersion	Limited or inconsistent	Limited or inconsistent	Promising; needs further validation

Figure 2

Bubble Heatmap of Design Logics by Educational Outcome and Effect



Challenge-based progression was the clearest pattern (E1, G4, H9, J6, J9) showing that escalating quests calibrated to the learner's level sustained flow and perceived usefulness, especially when difficulty was algorithmically adapted (G4, J6). These studies suggest that precise goal framing and immediate feedback, rather than the game "theme" itself, account for increased engagement.

Points and rewards also improved UX, chiefly by clarifying progress. In D2, F4, H8 and J9 the visibility of score tallies reduced learners' anxiety about their position in the course, while leaderboards (D2, G7, H9) fostered social presence. Nevertheless, H9 indicates that public ranking can undermine enjoyment for low-proficiency groups, implying that competitive displays require opt-out or cohort-based views.

Narrative framing and avatars produced more nuanced gains. When storylines were integrated with learning tasks (E1; avatar-rich H9) participants experienced deeper affective engagement. In opposition, studies where narrative was simply decorative (F4) did not report clear UX effects, emphasising that storytelling must scaffold, not obscure the pedagogy.

Personalisation logics acted as multipliers. Adaptive difficulty (B3, D2, G5, J4), learner-directed pathways (F2), and personalised feedback (H9, I8) collectively enhanced usability and relevance, specifically for novices who risk early abandonment. Remarkably, the highest composite UX scores arose in designs that layered gamified challenges with adaptive support (G5, J6), whereas single-feature deployments, such as quizzes without feedback (H7) or competitive points alone (H9), produced mixed outcomes. Overall, the evidence implies that optimal UX in ELT-oriented VLEs stems from synergistic combinations of challenge, clarity and personalisation rather than from any isolated mechanic.

Effects of Gamification on Motivation and Academic Performance (RQ2)

Gamified elements proved far more reliable in boosting learner motivation than in raising measurable achievement. Most studies (31/46) found at least some motivational benefit, and seven showed a statistically significant increase (e.g. D2, F2, J4, J6, J11, J13). Challenge-based designs were the most powerful driver (22 positive reports), followed by points (e.g. E1, G5, G10, H4, H9, I3), quizzes (e.g. D2, G4, H8), and game-based platforms (e.g. D2, F2, G10, H1). Where challenges were adaptive (G5, J6) or embedded in narrative quests (E1), learners described higher task value and sustained effort. By contrast, quiz-only implementations (H8) and static leaderboards sometimes yielded neutral motivation (G7), especially among low-proficiency cohorts.

Evidence was markedly mixed on academic performance. Only 15 studies detected a clear achievement improvement (e.g. D2, G5, F5, G10, H2, H9), while the modal pattern was "no significant difference" (31 studies). Positive effects clustered around designs that coupled gamified feedback with concept-aligned practice, for instance, adaptive Kahoot! drills (D2) or branching quests with formative checkpoints (J6). However, high-frequency challenge mechanics alone (19; mixed effect) rarely translated into higher test scores, suggesting that motivational lift does not automatically convert to learning outcomes without tight curricular integration and reflection opportunities.

The corpus indicates that gamification is a robust driver for engagement but a contingent one for achievement; its impact on grades depends on whether the mechanics are pedagogically aligned and supported by adaptive scaffolds rather than used as standalone motivators.

PL Strategies and Their Impact on Learning and Satisfaction (RQ3)

Personalisation was generally a stronger lever for learner satisfaction than for achievement, yet three tactics, adaptive feedback loops, learner-directed pathways, and fine-grained content tailoring, showed promise on both fronts. Adaptive difficulty appeared in two studies with positive motivational effects (B1, F5) and led to small but significant improvement in test scores and satisfaction. Learners expressed that tempo-matched tasks reduced frustration and improved control (B1).

Learner choice/path consistently maintained motivation (8; +; e.g. E2, F2, F4) and satisfaction (5; +; e.g. F2, F5), and improved academic achievement (4; +; e.g. F2, J6). Self-selection of sequence, especially when coupled with progress dashboards, fostered agency and deeper strategy use. Personalised content (6 satisfaction +, e.g. B3, F2; 4 performance +, e.g. F5, G5) and personalised feedback were valued for relevance and timeliness; both strategies correlated with higher post-test means in task-aligned vocabulary and grammar modules (H9, F5).

Profiles/preferences displayed limited evidence of positive effects on motivation (B3, F5) and academic performance (F5, G6), but more consistent effects on satisfaction (F5, G6, J9, J10). Self-paced tracking (G6, G10) and emerging AI-driven content adaptation (J2, J3, J4, J6) registered isolated but clear improvements in retention tests, albeit the small sample (≤ 4 studies each) cautions against over-generalisation.

Within PL studies, satisfaction gains slightly outnumbered those in performance (by a ratio of approximately 1.2:1), suggesting that students tend to value perceived personal relevance even when measurable learning improvements are less pronounced. Positive achievement effects surfaced primarily where personalisation was tightly coupled to formative feedback and curricular alignment; standalone preference toggles or static profiles rarely moved the performance needle. The evidence supports personalisation as

a reliable enhancer of perceived quality and a conditional contributor to learning, contingent on its depth of integration with pedagogical goals.

Challenges, Contradictions, and Methodological Gaps (RQ4)

Although the evidence base covers a wide range of contexts, three persistent fault lines (implementation hurdles, divergent findings, and design–method gaps) undermine the generalisability of results. Table 3 summarises these challenges and their categories, detailing the typical issues described, and the representative studies supporting them.

While the overall trends point to motivational and UX benefits, several contradictions emerged across the corpus. Table 4 illustrates three recurring tensions observed in the findings, each backed by specific studies. These contradictions highlight the nuanced impact of gamification and personalisation and reinforce the need for contextual sensitivity in design and evaluation.

Collectively, these gaps explain why motivational effects seldom translate into robust learning effects and why findings vary across cultural and infrastructural contexts. Future work must couple gamified and personalised designs with curricular alignment and teacher mediation to convert engagement into achievement. It should also adopt longer, mixed-method trials with validated UX instruments to capture sustained learning trajectories, and deliberately sample low-resource and younger cohorts to broaden ecological validity. By addressing these methodological shortfalls, forthcoming studies can produce more definitive guidance for designers and teachers seeking to combine gamification and personalisation in ELT-oriented VLEs.

In summary, the findings directly address the RQs of this SLR by identifying key design features (RQ1), analysing the impact of gamification and PL on learner motivation/engagement and academic outcomes (RQ2–RQ3), and highlighting common implementation challenges in VLEs (RQ4).

DISCUSSION

This section interprets this SLR's main findings, situating them within established theory and prior evidence, and then evaluates the methodological strengths and weaknesses that influence those patterns. It closes by outlining practical implications for ELT stakeholders and priority directions for future research.

Key Findings

This SLR reveals that pedagogical fit and interface quality, not the mere presence of gaming or adaptive features, de-

Table 3

Reported Implementation Challenges by Domain

Challenge Domain	Typical Issues	Representative Studies
Technical/usability	Platform instability, limited bandwidth, device restrictions, teachers' and students' digital training and literacy.	A1, D4, E2, F5, F6, G4, G8, G9, G12, H8, J1
Pedagogical alignment	Weak integration with syllabus, teacher workload, insufficient scaffolding	B2, E2, F4, F6, G1, G3, H1
Cultural/linguistic fit	Competitive mechanics clashing with collectivist norms, language-level mismatches, preference for face-to-face instruction.	B3, D4, F5
Sample & attrition	Small cohorts, voluntary dropout, uneven group sizes, need for family support	G12, H8, I2, J3
Measurement validity	Over-reliance on self-report scales, non-validated instruments	G5, I3, I5, J3
Short intervention span	Durations ≤ 4 weeks too brief to detect achievement gains	I2, J1, J3

Table 4

Contradictions in the Effects of Gamification and Personalisation Strategies

Contradiction	Description	Representative studies
Motivation ↑ vs. Performance ↔	Many studies reported high motivation but no significant improvement in academic performance, especially when gamification lacked pedagogical alignment.	H7 (strong motivation, weak test gains), J6 (both gains, due to strong adaptive scaffolding)
Leaderboards: engagement or anxiety?	Leaderboards boosted engagement through social presence, but sometimes harmed low-performing students' experience, particularly in collectivist contexts.	D2; G7 (positive), H9 (negative for low-level learners)
Narrative coherence vs. superficial appeal	Integrated storylines enhanced UX only when they supported the learning task; decorative narratives had no positive effect.	E1; I5 (positive), F4 (neutral)

termine learning impact on ELT-oriented VLEs. When challenges, points, or adaptive paths were embedded in clear interfaces that minimised extraneous cognitive load and delivered immediate, informative feedback (E2, G4, J7), learners perceived fluent navigation and higher task value. By contrast, the same mechanics led to a poor or neutral UX where apps were pedagogically weak (H4), or where basic usability was hampered by clunky navigation and distracting adverts (A1). This pattern aligns with Cognitive Load Theory (Sweller et al., 2011) and extends Dicheva et al.'s (2015) observation that usability lapses can nullify gamification benefits.

Motivational enhancements proved broad but fragile. Quizzes, points, and escalating challenges reliably triggered short-term behavioural engagement, confirming Self-Determination Theory's claim that clear goals and competence signals energise learners (Ryan & Deci, 2000). However, achievement outcomes emerged only when those mechanics were aligned with the curriculum and paired with formative feedback (D2, G5, J6). The recurring engagement–learning gap mirrors Looyestyn et al. (2017) and Koivisto & Hamari (2018) and suggests that extrinsic rewards alone rarely prompt deep processing.

The mixed appeal of social-comparison mechanics highlights the role of learner disposition. While leaderboards motivated some active users (D2, H9), this effect was not universal, with other learners showing disinterest (H9) or scepticism towards peer feedback (A1). This tension, where social features can both engage and alienate, reflects concerns noted by Antonaci et al. (2019) regarding their potential for negative affective outcomes. By contrast, well-integrated narratives achieved more consistent success, fostering immersive and supportive environments that focused learners on the task (E1, I5). These findings indicate that gamification's efficacy is contingent on an implementation that carefully balances learners' needs for autonomy, competence, and relatedness.

PL strategies proved the strongest dual impact when they combined adaptive feedback, learner choice, and tailored content. Such designs simultaneously satisfied autonomy and competence needs, leading to both higher satisfaction (F2, H9) and modest but significant performance results (B1, J6). Nevertheless, PL success depended on data quality, instructor mediation, and infrastructure; in low-resource contexts, adaptive algorithms often defaulted to "one-size-fits-all," muting their benefits.

Finally, contextual diversity (urban/rural schools, adolescent/adult learners, differing digital literacy) explains the diverse evidence. High-bandwidth and teacher-supported environments converted motivation into achievement, whereas self-paced or low-connectivity settings rarely did. Thus, no single mechanic guarantees success; meaningful impact arises only when interface clarity, adaptive scaffolds, and assessment feedback are consistent with learners' profiles and institutional realities.

Comparison with Prior Studies and Reviews

These results broadly confirm earlier gamification and PL syntheses while adding an ELT-specific perspective. As Dicheva et al. (2015) and Hamari et al. (2014) state, feedback-rich mechanics (quizzes, points, challenges) consistently increase behavioural engagement; nevertheless, echoing Sailer and Homner's (2020) meta-analysis, our data show that such increases in motivation rarely translate into academic improvement unless formative feedback and mastery thresholds are embedded.

Unlike most prior reviews, this SLR treated UX as an independent outcome. This revealed that interface clarity and cognitive-load management, rather than any single mechanic, drive positive UX, supporting the usability emphasis proposed by Seaborn and Fels (2015). For PL strategies, our findings align with recent meta-analytic evidence; adaptive difficulty and personalised content lead to the most reliable academic benefits (Fraulini et al., 2024), though their motivational value weakens when learners sense a loss of autonomy (Fong et al., 2019).

Finally, we corroborate Koivisto and Hamari's (2019) claim that context and implementation fidelity moderate outcomes: social-comparison features benefited high achievers but discouraged novices, a pattern also observed in Antonaci et al.'s (2019) review. Our ELT lens highlights that language-proficiency gaps intensify these divergences, pointing to a need for tiered or anonymous ranking systems in language classrooms.

Limitations

Despite following PRISMA 2020 guidelines, this review has several limitations. Grey literature was excluded, and searches were restricted to four databases covering 2015–2024, which could bias the evidence by favouring studies that are actually published, privileging those available in English, and missing research that takes longer to reach print.

Moreover, study coverage was uneven. Most investigations centred on higher-education contexts, with scant attention to secondary education or low-resource contexts. Geographically, research clustered in a few parts of Asia and Europe, leaving other areas, especially the Americas and Africa, underrepresented. Consequently, many of the design strategies discussed rest on assumptions such as stable broadband, individual devices, and competitive learning norms that seldom hold in those contexts; what works in well-resourced universities may be impractical or even counter-productive in rural secondary schools where con-

nectivity is deficient, devices are shared, and collaboration is valued over competition. Until these approaches are tested under such constraints, their broader applicability remains uncertain.

Methodological rigour was also limited: only two RCTs appeared among 46 studies; the rest relied on quasi-experimental or single-group designs with small convenience samples, a pattern prone to novelty effects and low external validity. Motivation and UX were often assessed with ad hoc, non-validated instruments. Since UX is inherently multidimensional, this reliance makes the UX findings exploratory and should temper any comparison with motivation or performance outcomes. The absence of validated UX tools thus reduces the reliability of conclusions. Similarly, academic results ranged from isolated vocabulary quizzes to full proficiency tests, hampering cross-study comparability. Finally, most studies gave only limited descriptions of how their gamification or personalisation elements were carried out, making it difficult to judge whether those features were delivered as intended.

Implications

Building on the evidence, the key implications are directed to three stakeholder groups. The emphasis is on how gamified and PL features are designed, integrated, and evaluated, rather than just being present. Educators should treat game mechanics and adaptive tools as scaffolds, intervening with formative feedback when dashboards reveal stagnation and using anonymised or tiered competition to protect novice confidence. Instructional designers must privilege usability over feature counts: mobile-first interfaces, offline caching, and transparent adaptive rules reduce cognitive load and allow teachers to override algorithmic decisions when needed.

Researchers can consolidate the evidence base through long-term, multi-site studies that couple validated UX and motivation scales with curriculum-aligned performance tests, paying particular attention to secondary schools in underrepresented regions such as Latin America and Africa. Reliable instruments, such as the User Engagement Scale, AttrakDiff, or the User Experience Questionnaire for UX, and the Intrinsic Motivation Inventory, MEEGA+, or the Motivated Strategies for Learning Questionnaire for motivation, should be combined with standardised language assessments (e.g., Cambridge Progress Tests, TOEFL Junior modules, or focused grammar cloze tasks). These measures need to be applied both formatively and summatively and clearly linked to specific domains (e.g., grammar accuracy versus communicative competence). Future work should also incorporate grey literature, preregister multi-site randomised controlled trials, and publish complete contextual and implementation details. Sharing open data and detailed implementation notes will shed light on how context, design quality, and pedagogical mediation transform engagement into lasting language gains, thereby enhancing the rigour and generalisability of the findings.

CONCLUSIONS

This SLR advances the field by showing that design quality and contextual fit, not only the presence of gamified or personalised features, determine learning impact in ELT-oriented VLEs. Treating UX as an independent outcome demonstrates that interface clarity and cognitive-load management support positive UX and mediate the translation of short-term motivation into durable achievement. It also clarifies why the same mechanics can benefit high achievers while alienating novices: cultural norms, language proficiency, and teacher mediation jointly shape learners' responses. The evidence suggests that meaningful gains arise only when adaptive scaffolds, formative feedback, and ethical competition are embedded in clear, low-load interfaces that respect local resource constraints.

Future research should move beyond short, single-site pilots to longitudinal, mixed-method studies that combine validated UX instruments, behavioural analytics, and curriculum-aligned performance tests. Particular attention is needed at the intersection of gamification and personalisation in low-resource settings, where mobile-first, offline-capable designs and teacher-controlled dashboards could foster more equitable benefits. Multi-site RCTs and well-reported quasi-experiments, when randomisation is unfeasible, will be critical for tracing whether initial motivational boosts mature into sustained language proficiency and learner autonomy.

For educational policy and practice, the findings establish three priorities. First, procurement guidelines should privilege usability audits, mobile resilience, and transparent adaptive engines over feature counts. Second, professional development must equip teachers to interpret learning analytics, modulate competitive elements, and integrate adaptive automation with scaffolded dialogue. Third, funding schemes and accreditation frameworks should promote rigorous trials in underrepresented regions and secondary classrooms, ensuring that evidence-based, context-sensitive VLEs become a realistic option for all English language learners.

ACKNOWLEDGEMENTS

Some parts of this manuscript were refined for language and style using ChatGPT, under the authors' direct supervision and review. All content was carefully checked and approved by the authors to ensure academic rigour and consistency with the purpose of the SLR.

DECLARATION OF COMPETING INTEREST

None declared.

AUTHORS' CONTRIBUTION

Myriam Tatiana Velarde Orozco: Conceptualization; Data curation; Formal analysis; Funding acquisition; Methodology; Project administration; Visualization; Writing – original draft; Writing – review & editing.

Bárbara Luisa De Benito Crosetti: Formal analysis; Investigation; Methodology; Resources; Software; Supervision; Writing – original draft.

REFERENCES

Al-Busaidi, K. A., & Al-Shihi, H. (2012). Key factors to instructors' satisfaction of learning management systems in blended learning. *Journal of Computing in Higher Education*, *24*(1), 18–39. https://doi.org/10.1007/s12528-011-9051-x

Ali, M., Wahab, I. B. A., Huri, H. Z., & Yusoff, M. S. (2024). Personalised learning in higher education for health sciences: a scoping review protocol. *Systematic Reviews*, *13*, Article 99. https://doi.org/10.1186/s13643-024-02478-4

Antonaci, A., Klemke, R., & Specht, M. (2019). The effects of gamification in online learning environments: A systematic literature review. *Informatics*, *6*(3), 32. https://doi.org/10.3390/informatics6030032

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*(4), 1675–1715. https://doi.org/10.1007/s10648-021-09615-8

Chaichumpa, S., Wicha, S., & Temdee, P. (2021). Personalized learning in a virtual learning environment using modification of objective distance. *Wireless Personal Communications*, *118*(3), 2055–2072. https://doi.org/10.1007/s11277-021-08126-7

Cho, I., Yeo, J., Hwang, G., & Yang, H. (2022). Impact of a virtual environment on learning effectiveness, motivation, cognitive load, and group self-efficacy of elementary school students in collaborative learning. *Educational Technology Research and Development*, *70*(6), 2145–2169. https://doi.org/10.1007/s11423-022-10159-z

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

Dicheva, D., Dichev, C., Agre, G., & Angelova, G. (2015). Gamification in education: A systematic mapping study. *Educational Technology & Society*, *18*(3), 75–88. http://www.jstor.org/stable/jeductechsoci.18.3.75

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

Fong, C. J., Patall, E. A., Vasquez, A. C., & Stautberg, S. (2019). A meta-analysis of negative feedback on intrinsic motivation. *Educational Psychology Review*, *31*(1), 121–162. https://doi.org/10.1007/s10648-018-9446-6

Fraulini, N. W., Marraffino, M. D., Garibaldi, A. E., Johnson, C. I., & Whitmer, D. E. (2024). Adaptive training instructional interventions: A meta-analysis. *Military Psychology*. Advance online publication. https://doi.org/10.1080/08995605.2024.2377884

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, *74*(1), 59–109. https://doi.org/10.3102/00346543074001059

Gevorgyan, S. (2024). The use of adaptive learning technologies in e-learning for inclusive education: A systematic review. *E-Learning Innovations Journal*, *2*(1), 90–107. https://doi.org/10.57125/ELIJ.2024.03.25.05

Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work? – A literature review of empirical studies on gamification. In *Proceedings of the 47th Hawaii International Conference on System Sciences* (pp. 3025–3034). IEEE. https://doi.org/10.1109/HICSS.2014.377

Helvich, J., Novak, L., Mikoska, P., & Hubalovsky, S. (2023). A Systematic Review of Gamification and Its Assessment in EFL Teaching. *International Journal of Computer-Assisted Language Learning and Teaching (IJCALLT)*, *13*(1), 1–21. https://doi.org/10.4018/IJCALLT.322394

Hollister, B., Nair, P., Hill-Lindsay, S., & Chukoskie, L. (2022). Engagement in online learning: Student attitudes and behaviour during COVID-19. *Frontiers in Education*, *7*, Article 851019. https://doi.org/10.3389/feduc.2022.851019

Hong, Q. N., Pluye, P., Fàbregues, S., Bartlett, G., Boardman, F., Cargo, M., Dagenais, P., Gagnon, M. P., Griffiths, F., Nicolau, B., O'Cathain, A., Rousseau, M. C., & Vedel, I. (2018). *Mixed Methods Appraisal Tool (MMAT), version 2018*. Canadian Intellectual Property Office, Industry Canada. http://mixedmethodsappraisaltoolpublic.pbworks.com/

Hung, H.-T., Yang, J.-C., Hwang, G.-J., Chu, H.-C., & Wang, C.-C. (2018). A scoping review of research on digital game-based language learning. *Computers & Education*, *126*, 89–104. https://doi.org/10.1016/j.compedu.2018.07.001

Klašnja-Milićević, A., Vesin, B., Ivanović, M., & Budimac, Z. (2011). E-learning personalization based on hybrid recommendation strategy and learning-style identification. *Computers & Education*, *56*(3), 885–899. https://doi.org/10.1016/j.compedu.2010.11.001

Koivisto, J., & Hamari, J. (2019). The rise of motivational information systems: A review of gamification research. *International Journal of Information Management*, *45*, 191–210. https://doi.org/10.1016/j.ijinfomgt.2018.10.013

Li, X., Xia, Q., Chu, S. K. W., & Yang, Y. (2022). Using gamification to facilitate students' self-regulation in e-learning: A case study on students' L2 English learning. *Sustainability*, *14*(12), 7008. https://doi.org/10.3390/su14127008

Looyestyn, J., Kernot, J., Boshoff, K., Maher, C., & Vandelanotte, C. (2017). Does gamification increase engagement with online programs? A systematic review. *PLOS ONE*, *12*(3), e0173403. https://doi.org/10.1371/journal.pone.0173403

Majuri, J., Koivisto, J., & Hamari, J. (2018). Gamification of education and learning: A review of empirical literature. In *Proceedings of the 2nd International GamiFIN Conference 2018* (pp. 11–19). CEUR-WS.org. http://ceur-ws.org/Vol-2186/paper2.pdf

Marín-Juarros, V. I. (2022). La revisión sistemática en la investigación en tecnología educativa: Observaciones y consejos. *Revista Interuniversitaria De Investigación En Tecnología Educativa*, (12), 62-79. https://doi.org/10.6018/riite.533231

Munn, Z., Peters, M. D. J., Stern, C., Tufanaru, C., McArthur, A., & Aromataris, E. (2018). Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Medical Research Methodology*, *18*, Article 143. https://doi.org/10.1186/s12874-018-0611-x

Nazempour, R., & Darabi, H. (2023). Personalized learning in virtual learning environments using students' behaviour analysis. *Education Sciences*, *13*(5), 457. https://doi.org/10.3390/educsci13050457

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, *372*, n71. https://doi.org/10.1136/bmj.n71

Pane, J. F. (2017). *How does personalized learning affect student achievement?* RAND Corporation. https://www.rand.org/pubs/research_briefs/RB9994.html

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Sailer, M., & Homner, L. (2020). The gamification of learning: A meta-analysis. *Educational Psychology Review*, *32*(1), 77–112. https://doi.org/10.1007/s10648-019-09498-w

Sanchis-Font, R., Castro-Bleda, M. J., González, J., Pla, F., & Hurtado, L. (2021). Cross-domain polarity models to evaluate user eXperience in E-learning. *Neural Processing Letters*, *53*(5), 3199–3215. https://doi.org/10.1007/s11063-020-10260-5

Seaborn, K., & Fels, D. I. (2015). Gamification in theory and action: A survey. *International Journal of Human-Computer Studies*, *74*, 14–31. https://doi.org/10.1016/j.ijhcs.2014.09.006

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L. A., & PRISMA-P Group. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: Elaboration and explanation. *BMJ*, *349*, g7647. https://doi.org/10.1136/bmj.g7647

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. Springer. https://doi.org/10.1007/978-1-4419-8126-4

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2012). Consumer acceptance and use of information technology: Extending the unified theory of acceptance and use of technology. *MIS Quarterly*, *36*(1), 157–178. https://doi.org/10.2307/41410412

Walkington, C., & Bernacki, M. L. (2020). Appraising research on personalized learning: Definitions, theoretical alignment, advancements, and future directions. *Journal of Research on Technology in Education*, *52*(3), 235–252. https://doi.org/10.1080/15391523.2020.1747757

Wichadee, S., & Pattanapichet, F. (2018). Enhancement of performance and motivation through application of digital games in an English language class. *Teaching English with Technology*, *18*(1), 77–92. https://eric.ed.gov/?id=EJ1170635

Zainuddin, Z., Chu, S. K. W., Shujahat, M., & Perera, C. J. (2020). The impact of gamification on learning and instruction: A systematic review of empirical evidence. *Educational Research Review*, *30*, Article 100326. https://doi.org/10.1016/j.edurev.2020.100326

Zhang, S., & Hasim, Z. (2023). Gamification in EFL/ESL instruction: A systematic review of empirical research. *Frontiers in Psychology*, *13*, Article 1030790. https://doi.org/10.3389/fpsyg.2022.1030790

APPENDIX A

DESCRIPTIVE matrix of included studies

This appendix offers a structured descriptive profile of the 46 studies reviewed in this SLR. The matrix summarises essential bibliographic and contextual information to support understanding of the research landscape and the diversity of implementations observed. It includes:

- Study identification and metadata (reference code, title, authors, publication year, journal, and source database).
- Geographical and educational context (country, educational level, and setting).
- Participant information (sample size and learner profile, including age, English proficiency, and digital competence when available).
- Methodological characteristics (study design and type).
- VLE description, where described.

Due to its extensive nature, the complete matrix is available in Figshare:

https://doi.org/10.6084/m9.figshare.29316077

APPENDIX B

MMAT ASSESSMENT OF Included Studies

This appendix presents the results of the quality appraisal conducted using the MMAT 2018. Each study was assessed according to its methodological category: qualitative, quantitative (randomised, non-randomised, or descriptive), or mixed methods.

Due to its extensive nature, the complete matrix is available in Figshare:

https://doi.org/10.6084/m9.figshare.29318828

APPENDIX C

Analytical matrix of included studies

This appendix offers a structured narrative summary of each study included in the review. For every article, it compiles key information on the VLE used, specific gamification and personalisation features (when specified), and the educational outcomes assessed such as motivation, engagement, UX, academic achievement, and learner satisfaction. It also captures author-reported findings, links between strategies and results, and implementation challenges.

Due to its extensive nature, the complete matrix is available in Figshare:

https://doi.org/10.6084/m9.figshare.29318840

APPENDIX D

Comparative Matrix of Design Logics and Educational Outcomes

	Design Logic	Motivation/ Engagement	UX	Academic Performance	Learner Satisfaction
Gamification	Avatars	4; +; H9	4; +; H9	3; +; H9	4; +; H9
	Badges/Achievements	1; +; F5	2; +; J9	1; +; F5	1; +; F5
	Challenges	22; +; H7	22; +; H7	19; +/-; H7	15; +/-; H7
	Competition	6; +; F5	5; +/-; F5	7; +/-; F5	6; +; G8
	Game-based platforms	7; +; H8	5; +; H8	7; +; H8	7; +; H8
	Leaderboards	3; +; H9	4; +; H9	3; +; H9	2; +; H9
	Levels	5; +; I2	4; +; I2	4; +; J1	5; +; H9
	Narrative/Story	2; +; E1	1; +; E1	1; -; E1	1; +; E1
	Points	9; +; J3	6; +; H8	10; +; F5	10; +; F5
	Quizzes	7; +/-; H8	7; +/-; H7	6; +/-; J9	8; +/-; H8
	Rewards	3; +; G10	4; +; G10	3; +; G10	3; +; G10
Personalisation	Adaptive difficulty	2; +; F5	2; +; F5	1; +; B1	1; +; B1
	Goals/Targets	2; +; F5	1; +; B1	1; +; B1	1; +; B1
	Learner choice/Path	8; +; F4	5; +; B1	4; +; F5	5; +; F5
	Personalised content	4; +; F5	6; +; F5	4; +; F5	6; +; F5
	Personalised feedback	6; +; H9	5; +; I8	5; +; B1	6; +; H9
	Profiles/Preferences	3; +; F5	3; +/-; F5	2; +; F5	4; +; F5
	Self-paced progress/ Tracking	2; +; G6	3; +; F5	1; +; G10	2; +; J10
	AI-driven content adaptation	4; +; J2	4; +; J3	4; +; J4	1; +; J2
	Augmented immersion	2; +; J1	2; +; J1	2; +; J1	2; +; J1