

Gamification And Human–Computer Interaction In Arabic Language Learning: A Systematic Review Of Engagement And Acquisition Outcomes

Moh. Ismail^{1,*} Alamuddin Ma'ruf²

¹Universitas Darussalam Gontor, Indonesia

²University of The Holy Qur'an and Islamic Sciences, Sudan

Received: 18 April 2026

Accepted: 5 May 2026

Published: 30 June 2026

ismail@unida.gontor.ac.id

alamuddin@quran-unv.edu.sd

*Correspondence: ismail@unida.gontor.ac.id

DOI: <https://doi.org/xxxx>

Abstract

Gamification and human–computer interaction (HCI) have emerged as consequential research domains in language education technology, yet their application to Arabic language learning remains systematically underexplored. This systematic review synthesises empirical evidence on the effects of gamification and HCI design principles on engagement and language acquisition outcomes among learners of Arabic as a second or foreign language. Following PRISMA guidelines, 32 peer-reviewed studies published between 2019 and 2024 were selected from four major databases: ERIC, Web of Science, Scopus, and the ACM Digital Library. Findings indicate that point-based reward systems, narrative scaffolding, and adaptive interface design yield statistically significant improvements in vocabulary retention and oral production when aligned with the phonological and morphological demands of Arabic. Leaderboard and competitive mechanics show mixed effects, contingent on learner motivation profiles and cultural context. Critically, most existing platforms demonstrate limited linguistic responsiveness to Arabic's diglossic character, script directionality, and non-concatenative morphology—constraining both engagement sustainability and acquisitional depth. The review proposes an Arabicised HCI–Gamification (AHG) design framework comprising four interdependent principles: Linguistic Anchoring, Motivational Calibration, Interface Fidelity, and Cultural Situatedness. Implications are discussed for instructional designers, Arabic curriculum developers, and EdTech researchers seeking theoretically grounded, culturally responsive approaches to interactive Arabic language learning.

Keywords: gamification, human–computer interaction, Arabic language learning, educational technology, learner engagement

Introduction

The integration of game-based mechanics into educational software—broadly termed gamification—has attracted sustained scholarly attention over the past decade as a strategy for enhancing learner motivation, persistence, and achievement in digital learning environments (Hamari et al., 2014; Dichev & Dicheva, 2017). Since Deterding et al. (2011) formalised the concept as the use of game design elements in non-game contexts, researchers have documented its application across diverse educational domains, from mathematics and science education to professional training and health literacy. In language learning, gamification has become particularly prominent: platforms such as Duolingo, Babbel, and Busuu have deployed badges, streaks, leaderboards, and progress bars at scale, generating substantial learner uptake and stimulating parallel academic inquiry into the conditions under which game mechanics translate into durable language acquisition (Reinhardt, 2019; Dehghanzadeh et al., 2021).

Parallel developments in human–computer interaction (HCI) research have deepened understanding of how interface affordances, feedback design, and interactive task structures mediate learner engagement and cognitive load in computer-assisted language learning (CALL) contexts (Heift & Schulze, 2007; Chapelle, 2021). The usability, responsiveness, and visual architecture of a learning interface are not peripheral to learning outcomes—they directly shape the quality and quantity of learner engagement with linguistic content. When an interface imposes unnecessary navigational complexity, produces ambiguous feedback signals, or misrepresents the target language's orthographic conventions, cognitive resources are diverted from the actual learning task (Mayer, 2021; Sweller, 1988). Together, gamification and HCI represent a convergent design paradigm for language education technology with considerable, though unevenly realised, potential.

Within the global landscape of language learning technology, Arabic occupies a structurally anomalous position. Although Arabic ranks among the world's most widely spoken languages—with over 400 million native speakers and an expanding global learner population that includes millions of heritage speakers, students of Islamic studies, and professionals across government, media, and international organisations—the Arabic CALL ecosystem lags measurably behind comparable infrastructure for languages such as English, Mandarin, or Spanish (Al-Hamad & Al-Shara, 2022; Alosch, 2023). This disparity is compounded by Arabic's distinctive linguistic architecture: its diglossic configuration, in which Modern Standard Arabic (MSA) and a heterogeneous array of regional dialects coexist in socially differentiated functional domains, creates a learning environment of unusual complexity (Ferguson, 1959; Versteegh, 2014). Its non-concatenative root-and-pattern morphology—in which meaning is encoded through the interweaving of consonantal roots and vowel-pattern templates rather than linear affixation—its right-to-left script directionality, and its phonological inventory including pharyngeal and uvular consonants absent from most learners' first language (L1) systems all impose design demands on CALL interfaces that general-purpose language platforms rarely address (Watson, 2002; Holes, 2004).

Empirical literature on gamification in Arabic language learning has grown modestly but unevenly. Studies have examined point systems, leaderboards, badges, and narrative game structures primarily in vocabulary acquisition contexts (Hassan & Ibrahim, 2024; Al-Otaibi, 2022), with comparatively limited attention to listening, speaking, reading, or integrated skills. HCI research focused on Arabic interfaces has similarly concentrated on specific design affordances—virtual keyboards, vowelisation systems, root-based search mechanisms—without systematic synthesis with gamification literature or second language acquisition (SLA) theory (Al-Tamimi & Kaur, 2021). A critical gap therefore exists: no integrated review has examined how gamification mechanics and HCI design principles interact with the linguistic

specificity of Arabic, nor evaluated the degree to which current platforms are responsive to the unique challenges that Arabic poses to both learners and interface designers. The absence of such synthesis constrains both research design and instructional practice, leaving developers without a theoretically grounded framework for making design decisions and researchers without a consolidated evidence base from which to build.

This systematic review addresses that gap. It synthesises peer-reviewed empirical evidence on the effects of gamification and HCI design on engagement and acquisition outcomes in Arabic language learning (Ismail et al., 2023), identifies patterns in current practice, surfaces design limitations attributable to insufficient linguistic responsiveness, and proposes the Arabicised HCI–Gamification (AHG) Framework as a principled basis for future research and development. The review is guided by three research questions: (1) What gamification mechanics and HCI design features have been empirically investigated in Arabic language learning contexts? (2) What are the demonstrated effects of these features on learner engagement and language acquisition outcomes? (3) What Arabic-specific linguistic and cultural variables moderate these effects, and how adequately are they addressed in existing platforms and studies?

2. Theoretical Framework

2.1. *Gamification and Self-Determination Theory*

The theoretical basis for understanding how gamification affects motivation in language learning is most coherently supplied by Self-Determination Theory (SDT; Deci & Ryan, 1985, 2000). SDT posits that intrinsic motivation—the primary driver of sustained learning engagement—is supported by the fulfilment of three basic psychological needs: autonomy (perceived control over one's learning trajectory), competence (the experience of effectiveness and progressive mastery), and relatedness (a sense of meaningful connection with others). Gamification mechanics exert motivational influence to the extent that they satisfy or thwart these needs. Point systems and progress bars may reinforce perceived competence by making incremental learning gains visible; collaborative game structures and social forums may enhance relatedness; learner-chosen difficulty levels and branching narrative choices may nurture a sense of autonomy (Mekler et al., 2017). The SDT framework thus predicts that gamification is not uniformly beneficial: its effects depend critically on which needs the mechanics activate and whether they do so in ways that feel self-determined rather than externally controlled.

Critically, the relationship between gamification and intrinsic motivation is not guaranteed. Domínguez et al. (2013) found that while gamified instruction produced short-term performance gains, it did not consistently enhance intrinsic motivation, and competitive mechanics—particularly public leaderboards—risked undermining the sense of autonomy and relatedness in learners who perceived themselves as low-performing relative to peers. Ryan and Deci (2020) further clarify that external rewards, including points and badges, can crowd out intrinsic motivation if they become the primary rationale for engagement—a dynamic particularly likely when rewards are contingent, expected, and salient. For Arabic learners, many of whom approach the language with high instrumental motivation but significant anxiety about its script, phonological unfamiliarity, and morphological complexity (Oxford et al., 2020), the SDT lens predicts that gamification design must be carefully calibrated: mechanics that amplify competitive pressure or render failure conspicuous may be particularly counterproductive, whereas mechanics that scaffold progressive mastery of concrete Arabic

sub-skills—phoneme discrimination, script recognition, root identification—may genuinely support competence-need satisfaction.

HCI Design Principles and Cognitive Load Theory

Human–computer interaction in language learning contexts is principally governed by the interplay between interface affordances and cognitive load. Sweller's (1988) Cognitive Load Theory (CLT) distinguishes between intrinsic load (inherent to the complexity of the learning content itself), extraneous load (imposed by the design of the learning interface), and germane load (cognitive effort directed toward schema formation and long-term memory consolidation). Effective HCI design minimises extraneous load—reducing the cognitive cost of navigation, text input, feedback interpretation, and visual parsing—so that available cognitive resources are directed toward the target language content rather than toward the interface itself (Mayer, 2021). A corollary principle from Mayer's (2021) Cognitive Theory of Multimedia Learning is the coherence effect: removing extraneous material from a learning interface improves retention precisely because it reduces the cognitive competition for limited working memory capacity.

For Arabic, CLT implications are substantially amplified. Learners must simultaneously manage the right-to-left reading directionality, the unvowelled script standard in adult literacy contexts (rasm), the morphological parsing demands of root-pattern recognition, and—in listening contexts—the diglossic gap between instructed MSA and the dialectal Arabic encountered in authentic communicative situations. This constellation of intrinsic loads, which most CALL interfaces designed with left-to-right scripts in mind inadvertently exacerbate through poor localisation, creates a fundamentally different cognitive learning environment than that assumed by general HCI usability research (Al-Tamimi & Kaur, 2021). The mixed-directionality problem—Arabic target language content embedded within a predominantly left-to-right interface architecture—imposes additional extraneous load by requiring learners to continually recalibrate their reading direction, an attentional cost that compounds across extended learning sessions. HCI design for Arabic language learning must therefore address not merely generic usability principles but the Arabic-specific extraneous loads generated by script rendering, input method design (Arabic keyboards, diacritical marking tools), and the visual presentation of morphologically related word families.

Flow Theory and Engagement

Csikszentmihalyi's (1990) Flow Theory offers a complementary framework for understanding the conditions under which gamification sustains deep engagement over time. Flow—the state of complete absorption in a challenging but manageable activity, characterised by focused concentration, loss of self-consciousness, and intrinsic enjoyment—is most reliably elicited when task challenge is precisely calibrated to learner skill level. When challenge substantially exceeds skill, anxiety disrupts engagement; when skill substantially exceeds challenge, boredom follows. Gamified language learning environments that incorporate adaptive difficulty mechanisms are, in principle, flow-generative: by continuously adjusting challenge to competence, they maintain the narrow channel in which sustained engagement and deep processing are simultaneously possible (Reinhardt, 2019). Flow states have been associated empirically with superior memory consolidation and transfer, making flow-generative design not merely motivationally desirable but acquisitionally consequential.

For Arabic language learning specifically, the challenge–skill calibration problem is among the most demanding of any widely taught language. The U.S. Foreign Service Institute (FSI, 2023) estimates approximately 2,200 hours of instruction for a native English speaker to

reach professional working proficiency in Arabic—roughly four times the investment required for Romance or Germanic languages. This steep learning curve creates a wide skill distribution among Arabic learners even within ostensibly homogeneous instructional groups, and gamification platforms must therefore model individual proficiency with considerably greater granularity than is typical for less challenging languages. Platforms that apply a single difficulty trajectory to all learners within a given CEFR level risk disrupting flow states for substantial proportions of their users—either by failing to provide sufficient challenge for faster learners or by maintaining challenge levels that generate sustained anxiety in slower ones.

SLA Perspectives on Input, Output, and Interaction

Second language acquisition theory provides a further theoretical lens through which the acquisitional implications of gamified Arabic CALL can be evaluated. Krashen's (1985) Input Hypothesis remains influential: acquisition depends on comprehensible input pitched slightly above the learner's current competence level—the $i+1$ principle—and gamification contexts can either facilitate or impede access to appropriately calibrated input depending on their design. Swain's (1985) Output Hypothesis extends this account by arguing that comprehensible output—being pushed to produce language at the edge of one's competence—also drives acquisition through processes of noticing, hypothesis testing, and metalinguistic reflection. Gamification mechanics that require learners to produce Arabic under time pressure, respond to comprehension checks, or engage in dialogue with AI conversational agents may operationalise these processes in accessible and motivationally engaging ways.

Long's (1996) Interaction Hypothesis adds a third dimension: negotiation of meaning between speakers—particularly in contexts where communication breakdown prompts repair, clarification, and reformulation—is among the most powerful drivers of acquisition. In gamified Arabic CALL, interaction is typically simulated through branching dialogue systems, adaptive feedback loops, or AI-powered conversational agents rather than human-to-human exchange. (Ismail, 2019) The degree to which such simulated interaction approximates the acquisitional properties of genuine negotiation of meaning remains an open empirical question, though evidence from intelligent tutoring systems suggests that well-designed AI-mediated interaction can produce learning outcomes approaching those of human tutoring in specific skill domains (VanLehn, 2011).

Method

Research Design

This study employed a systematic review methodology following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (Page et al., 2021). Systematic review was selected as the appropriate design given the study's objectives: to synthesise existing empirical evidence across diverse methodological traditions, identify convergent and divergent patterns in findings and design approaches, and establish a principled basis for proposing a novel conceptual framework. No primary data collection was undertaken, and the study did not involve direct interaction with human participants. The review protocol was registered prior to data extraction in order to ensure methodological transparency and limit post-hoc modifications to the analytical framework.

Inclusion and Exclusion Criteria

Studies were included if they: (a) investigated gamification mechanics, HCI design features, or their interaction in a language learning context; (b) reported empirical data, whether quantitative, qualitative, or mixed methods; (c) addressed Arabic as the target language of learning or included Arabic-specific interface analysis as a primary concern; (d) were published in peer-reviewed journals or conference proceedings between January 2019 and December 2024; and (e) were available in English. Studies were excluded if they: (a) examined gamification in non-language academic subjects without direct connection to language acquisition outcomes; (b) were theoretical or opinion pieces without empirical data; (c) addressed Arabic as a medium of instruction in non-language subjects rather than as a target language; or (d) reported interventions with fewer than ten participants, given inadequate statistical power for meaningful effect size estimation.

Search Procedures

Electronic database searches were conducted across four databases: ERIC, Web of Science, Scopus, and the ACM Digital Library. Search strings combined controlled vocabulary and free-text terms across three concept clusters: (1) gamification OR game-based learning OR serious games OR digital game OR points OR badges OR leaderboards; (2) human–computer interaction OR HCI OR interface design OR CALL OR computer-assisted language learning OR usability; and (3) Arabic language learning OR Arabic as a foreign language OR Arabic as a second language OR Arabic CALL OR Arabic EdTech. Boolean operators and database-specific field tags were applied to optimise recall and precision. Initial searches were conducted in April 2024 and comprehensively updated in November 2024 to capture recently published studies. Reference lists of all included studies were manually screened for additional eligible sources not captured through database searches.

Study Selection and Data Extraction

Title and abstract screening was conducted independently by two reviewers using Rayyan systematic review software, with disagreements resolved through structured discussion and, where necessary, full-text consultation. Full-text review of potentially eligible studies was subsequently performed by both reviewers, applying the inclusion and exclusion criteria to each candidate study. Data were extracted using a standardised, pre-piloted form capturing: study design and methodological approach, participant characteristics (number, proficiency level, L1, instructional context), gamification mechanics and HCI design features investigated, outcome measures and instruments, key statistical or thematic findings, and Arabic-specific design considerations reported. Methodological quality was assessed using the Mixed Methods Appraisal Tool (MMAT; Hong et al., 2018), which enables quality appraisal across quantitative, qualitative, and mixed methods study types within a single framework. Following full-text screening, 32 studies met all inclusion criteria and were retained for synthesis. All procedures were conducted in compliance with relevant ethical standards for secondary research synthesis; no ethical clearance for human participants was required.

Findings and Discussion

Gamification Mechanics in Arabic Language Learning

Of the 32 included studies, the most frequently investigated gamification mechanics were point systems and experience points ($n = 21$), badges and achievement markers ($n = 19$), leaderboards ($n = 17$), narrative structures and storylines ($n = 12$), and adaptive difficulty

systems (n = 9). Progress bars, virtual currency systems, and time-pressured challenge modes were investigated in smaller subsets of studies. This distribution broadly mirrors the gamification mechanics most commonly deployed in general CALL contexts (Dehghanzadeh et al., 2021), suggesting that Arabic language platforms have largely replicated generic gamification design patterns rather than developing mechanics tailored to Arabic's distinctive acquisitional demands. Notably, the representation of Arabic-specific adaptations within each mechanic category—such as root-based progression maps, calligraphic reward visuals, or culturally grounded narrative settings—was sparse: fewer than a quarter of studies overall reported any Arabic-responsive gamification design feature, indicating a substantial design gap between available gamification approaches and the linguistic requirements of Arabic instruction.

Vocabulary acquisition was the most common outcome measured across included studies (n = 24), followed by reading comprehension (n = 13), writing accuracy (n = 9), and speaking or oral production (n = 7). Listening comprehension, despite its theoretical prominence in SLA and its particular challenge for Arabic learners navigating the diglossic landscape, featured in only five studies—a finding consistent with the broader pattern of relative neglect of listening skill in the Arabic CALL literature (Vandergrift & Goh, 2012; Al-Khatib & El-Imam, 2021) and one that highlights a persistent and educationally consequential research lacuna.

Table 1. Distribution of Gamification Mechanics and Outcome Measures Across Included Studies (N = 32)

Gamification Mechanic	Studies (n)	Primary Outcome Measured	Arabic-Specific Adaptation
Point systems / XP	21	Vocabulary retention	8 (38%)
Badges / Achievements	19	Vocabulary retention, reading	6 (32%)
Leaderboards	17	Vocabulary, motivation	4 (24%)
Narrative structures	12	Reading, writing, motivation	7 (58%)
Adaptive difficulty	9	All skills combined	5 (56%)
Virtual currency	6	Vocabulary, engagement	2 (33%)
Time-pressure challenges	5	Listening, speaking	2 (40%)

Note. Arabic-Specific Adaptation = proportion of studies in that mechanic category that reported any Arabic-responsive design feature.

The relatively higher rate of Arabic-specific adaptation in narrative structures (58%) and adaptive difficulty systems (56%) compared with point systems and leaderboards (24–38%) is theoretically coherent and practically significant. Narrative scaffolding creates contextualised opportunities for embedding Arabic-specific content—culturally grounded storylines, diglossic dialogue sequences, script-integrated world-building in which Arabic calligraphy and typographic conventions are woven into the game visual language—whereas the abstract reward logic of points and badges is less inherently language-specific and therefore less likely to prompt designers to consider Arabic's particular demands. Adaptive difficulty systems that

explicitly modelled phonological and morphological sub-competencies—distinguishing, for example, between a learner's phoneme discrimination accuracy for pharyngeal consonants and their root-pattern recognition speed—similarly enabled more language-responsive calibration than those using generic CEFR-level proxies, which collapse important Arabic-specific variation across skill sub-components.

Effects on Engagement and Language Acquisition

Across the 32 studies, gamification interventions were associated with statistically significant improvements in learner engagement metrics—including time-on-task, self-reported motivation, session frequency, and completion rates—in 26 of 32 studies (81%). This overall positive pattern must, however, be interpreted with methodological caution: the majority of included studies relied on self-report measures of motivation collected immediately post-intervention, with comparatively few studies assessing sustained engagement over periods exceeding four weeks ($n = 9$) or employing objective behavioural measures of engagement such as session logs, interaction traces, or eye-tracking data ($n = 7$). The 81% positive engagement finding likely overstates the long-term motivational benefits of gamification, given well-documented novelty effects in EdTech research: learners tend to exhibit elevated engagement with any new instructional technology during initial exposure, with effects attenuating as novelty diminishes (Reinhardt, 2019). Furthermore, the measurement of engagement as a construct varied substantially across studies, with some operationalising it as total time on platform, others as self-reported willingness to continue learning, and still others through behavioural proxies such as voluntary task repetition or unsolicited use outside scheduled instructional sessions. This measurement heterogeneity means that the 81% figure aggregates findings that are not strictly comparable, and should be interpreted as a directional indicator of positive engagement effects rather than a precise meta-analytic estimate. Future research in Arabic CALL gamification would benefit considerably from the adoption of standardised, multi-component engagement measurement frameworks that distinguish behavioural, cognitive, and emotional engagement dimensions (Fredricks et al., 2004).

Effects on language acquisition outcomes were more differentiated and, on balance, more modest than engagement findings alone might suggest. Point systems and badge mechanics showed significant effects on vocabulary retention in 17 of 21 studies (81%), with effect sizes ranging from small to moderate (Cohen's $d = 0.31$ – 0.72). However, transfer to authentic communicative tasks—assessed through spoken production, writing tasks, or real-world comprehension activities rather than isolated vocabulary recall measures—was evaluated in only four studies, and the evidence for retention beyond 30 days was limited to three studies, only two of which reported sustained effects. This pattern raises questions about whether point-and-badge mechanics, which are highly effective at driving repeated exposure to target vocabulary items, also support the deeper processing—contextual integration, morphological network building, pragmatic understanding—that Arabic vocabulary acquisition theory suggests is necessary for durable and transferable learning (Holes, 2004; Nation & Newton, 2009).

Leaderboard mechanics produced the most inconsistent pattern of findings across included studies. Positive effects on engagement and vocabulary performance were reported in competitive learner populations and in contexts where learners perceived themselves as high-performing relative to peers (Al-Otaibi, 2022; Shadiev et al., 2023). These positive effects were, however, systematically offset in three studies by findings of elevated anxiety, reduced session persistence, and lower self-efficacy among learners with low baseline proficiency or collectivist cultural orientations—precisely the learner profile that characterises many

institutional Arabic language learning contexts, particularly in Southeast Asian, South Asian, and Sub-Saharan African settings where Arabic is studied for religious as well as linguistic purposes. This pattern closely mirrors SDT's predictions about the relatedness-undermining and autonomy-threatening potential of social comparison mechanics (Mekler et al., 2017), and suggests that leaderboard deployment in Arabic CALL contexts requires careful contextual calibration rather than default adoption.

Narrative gamification structures yielded the strongest and most consistent evidence for integrated skill development across the included studies. Of the 12 studies investigating narrative mechanics, nine reported improvements across at least two language skill areas simultaneously, and five demonstrated evidence of transfer to spontaneous communication tasks beyond the immediate instructional context. This pattern suggests that contextualised, meaning-bearing game structures—in which learners engage with authentic Arabic cultural content as part of a sustained game narrative, encountering target language items in semantically and pragmatically coherent contexts—may offer superior acquisitional affordances compared with decontextualised drill-and-reward mechanics. This is consistent with embodied cognition perspectives on language learning (Kramsch, 2009), with Nation and Newton's (2009) four-strands model's emphasis on meaning-focused input as the primary driver of vocabulary acquisition, and with Long's (1996) Interaction Hypothesis in so far as narrative structures that involve problem-solving dialogues simulate meaning-negotiation processes.

HCI Design and Arabic Linguistic Variables

The HCI analysis of included studies revealed a systematic and educationally consequential gap between interface design sophistication and Arabic linguistic responsiveness. Twenty-four of the 32 included studies reported platforms with standard Latin-script interface architectures adapted for Arabic through surface-level localisation—primarily through Arabic font substitution and right-to-left text alignment adjustments—rather than through foundational HCI redesign attentive to Arabic's morphological, phonological, and grapho-tactile architecture. This distinction between localisation and foundational linguistic responsiveness is pedagogically critical. A platform that renders Arabic text correctly but presents vocabulary items as isolated orthographic strings—without root-family grouping, morphological annotation, diacritical vowelisation support, or visual cues to related word forms—imposes substantial extraneous cognitive load on learners attempting to build the morphological knowledge networks that Arabic vocabulary acquisition requires (Holes, 2004; Mayer, 2021). Such a platform may feel Arabic in its surface presentation while remaining cognitively structured around assumptions appropriate for morphologically simpler languages.

Arabic script directionality created measurable usability challenges across multiple studies that employed eye-tracking, interaction logging, or error-rate analysis. Bouhlal et al. (2022) found that mixed-directionality interfaces—in which Arabic target language items were presented in right-to-left orientation within a broader left-to-right interface architecture—produced significantly elevated error rates on reading tasks and longer task completion times compared with fully right-to-left interface designs. The magnitude of this effect increased with text density and decreased with learner experience, suggesting that interface directionality represents an extraneous load that partially but incompletely habituates over extended use. Al-Tamimi and Kaur (2021) similarly documented that Arabic learners rated mixed-directionality interfaces as significantly less usable and more cognitively effortful than Arabic-first designs, even when both interface types presented identical linguistic content. These findings confirm

that directionality is not a cosmetic feature of interface design but a cognitive infrastructure decision with direct consequences for learning efficiency and accessibility.

Diglossia remained the most underserved and, arguably, the most consequential Arabic linguistic variable across the HCI literature. Only six of the 32 included studies explicitly addressed the MSA/dialect distinction in their interface design or research findings, despite the well-documented discrepancy between the Arabic learners encounter in formal instruction and standardised testing (MSA) and the Arabic they are exposed to in authentic communicative contexts (regional dialects; Ferguson, 1959; Versteegh, 2014). Platforms that presented all game audio exclusively in MSA were rated by learners as less authentic and produced lower motivation scores in three studies that included user satisfaction measures—a finding that is educationally significant given the motivational role of perceived relevance in sustaining language learning engagement. Conversely, platforms that attempted to incorporate dialectal variation without pedagogical framing—presenting Egyptian colloquial Arabic as equivalent to MSA without contextual scaffolding—generated learner confusion and reduced self-efficacy in one quasi-experimental study. These findings collectively suggest that addressing diglossia in Arabic CALL HCI design requires not merely including dialectal content but providing learners with a coherent pedagogical framework for understanding the functional relationship between MSA and the dialect(s) represented.

Table 2. Arabic Linguistic Variables and Their Treatment in HCI Design Across Included Studies

Arabic Linguistic Variable	Studies Addressing It (n)	Adequate HCI Treatment (n)	Common Design Gap Identified
Script directionality (RTL)	28	14	Mixed-directionality interfaces increase extraneous load
Morphological root-pattern system	16	7	Vocabulary items presented in isolation; no root family grouping
Diacritical vowelisation (harakat)	18	9	Missing vowel marks reduce decodability for beginners
Diglossia (MSA vs. dialects)	6	3	Exclusive MSA input misaligns with authentic communication goals
Phoneme inventory (pharyngeals, etc.)	11	5	ASR-based pronunciation feedback trained on non-Arabic data

Note. RTL = Right-to-left; MSA = Modern Standard Arabic; ASR = Automatic Speech Recognition.

The Arabised HCI–Gamification (AHG) Framework

Synthesising the theoretical frameworks surveyed above and the empirical patterns identified across the 32 included studies, this paper proposes the Arabised HCI–Gamification (AHG) Framework as a design and evaluation lens for gamified Arabic language learning platforms. The framework is responsive to both the opportunities and the limitations identified in the current evidence base: it takes seriously the demonstrated potential of well-designed gamification and HCI to enhance Arabic learner engagement and skill development, while foregrounding the Arabic-specific linguistic, cultural, and pedagogical variables that existing generic frameworks inadequately address. The AHG Framework does not constitute a rejection of gamification or HCI innovation in Arabic language learning; rather, it constitutes a demand that such innovation be linguistically informed and pedagogically accountable. Generic gamification frameworks such as the Octalysis model (Chou, 2019) or the MDA (Mechanics, Dynamics, Aesthetics) framework (Hunicke et al., 2004) offer valuable structural scaffolding for game design thinking but are silent on the language-specific dimensions that Arabic uniquely demands. The AHG Framework fills that silence. It does so not by replacing existing gamification theory but by embedding it within an Arabic language acquisition context, translating abstract design principles into specific, actionable criteria that account for diglossia, morphological architecture, script directionality, and the cultural diversity of global Arabic learner communities. The AHG Framework comprises four interdependent design principles: Linguistic Anchoring, Motivational Calibration, Interface Fidelity, and Cultural Situatedness.

Linguistic Anchoring requires that gamification mechanics be designed to structurally reinforce, rather than abstract away from, Arabic's morphological and phonological architecture. Root-based progression maps—in which learners unlock vocabulary clusters organised around Arabic trilateral and quadrilateral roots, earning experience points not for isolated word memorisation but for demonstrating productive understanding of root-pattern families—exemplify this principle. The gamification mechanic (progression and unlocking) is in this design structurally homologous with the acquisitional task (building root-pattern morphological networks): the activity that earns rewards is the activity that advances Arabic morphological competence. By contrast, mechanics that treat vocabulary items as atomistic, morphologically isolated units—rewarding the correct identification of /mudarris/ without reinforcing its derivational relationship to /darasa/, /madrasah/, and /dirāsah/—fail the Linguistic Anchoring criterion regardless of their visual engagement.

Motivational Calibration applies SDT's insight that motivation is need-specific and context-dependent: gamification designers must diagnose which motivational needs are most salient for their target learner population and select mechanics accordingly, rather than defaulting to the competitive and extrinsic reward structures that dominate popular commercial language-learning applications. For Arabic learners in formal institutional contexts—university students of Arabic, madrasa learners, heritage-language maintainers—where instrumental motivation is typically high but anxiety about script and phonological difficulty is also elevated, mechanics that scaffold a sense of growing competence through transparent progress visualisation, mastery badges tied to specific Arabic sub-skills, and skill tree branching that makes learners' expanding capabilities visible are likely to be more motivationally sustaining than public leaderboards or time-pressure challenges (Oxford et al., 2020; Ryan & Deci, 2020). For self-directed adult learners with high autonomy motivation, by contrast, platform designs that offer meaningful choices about learning pathways, dialect focus, and content domains may be more motivationally responsive.

Interface Fidelity requires that HCI design treat Arabic's grapho-phonological system as a primary rather than secondary design variable—as a fundamental design constraint from

which interface architecture should be derived, not a localisation requirement that can be addressed through post-hoc font substitution. Fully right-to-left interface architecture, robust diacritical display and input controls, root-family vocabulary grouping as a default rather than an advanced feature, morphologically informed feedback systems that present error corrections in the context of root-pattern families, and ASR-based pronunciation feedback trained specifically on Arabic phonological data should be foundational features of Arabic CALL platforms. The extraneous load findings reviewed above confirm that interface fidelity has direct measurable consequences for acquisition efficiency: learners operating on interfaces that misrepresent Arabic's script logic must allocate working memory resources to compensating for interface deficiencies rather than to processing linguistic content.

Cultural Situatedness requires that narrative game structures, visual aesthetics, character representations, and audio environments reflect the sociolinguistic and cultural diversity of contexts in which Arabic is actually learned and used. This principle extends beyond cultural sensitivity to encompass the acquisitional logic of embodied cognition theory (Kramsch, 2009; Lakoff & Johnson, 1999): meaning construction in language learning is not culturally neutral but is fundamentally embedded in the cultural schemas, pragmatic conventions, and social contexts that give language its communicative force. A gamified Arabic learning environment whose narratives are culturally thin—generic adventure scenarios with Arabic text overlaid—provides learners with linguistic input without the cultural embedding that facilitates deep semantic processing and pragmatic competence. Cultural Situatedness therefore requires authentic dialectal variation in game audio (calibrated to the communicative goals of the learner population), storylines that draw meaningfully on Arabic literary, oral, and everyday cultural traditions, and character and setting design that reflects the geographic and cultural diversity of Arabic-speaking communities rather than homogenised stereotypes.

The four AHG principles are mutually constraining rather than independently applicable: a platform achieving high Interface Fidelity but misapplying Motivational Calibration will produce a linguistically accurate but pedagogically demotivating experience; a platform with strong Cultural Situatedness but weak Linguistic Anchoring will engage learners in authentic Arabic cultural content without directing that engagement toward the morphological and phonological processing tasks that Arabic acquisition demands. The framework is intended to function as an integrative design philosophy and evaluative lens, applicable both to the assessment of existing Arabic EdTech platforms and to the specification of future gamified Arabic CALL designs. Its ultimate empirical validation depends on experimental and quasi-experimental research that deploys, tests, and iteratively refines AHG-informed designs across diverse Arabic learning populations—a research agenda that the current review's findings both motivate and partially equip.

Implications for Arabic Language EdTech Practice

The findings of this review carry several practical implications for developers, instructors, and researchers working within the Arabic EdTech ecosystem. For platform developers, the most immediate implication is that Arabic cannot be adequately served by adapting existing multilingual platforms through superficial localisation. The evidence reviewed here consistently demonstrates that interfaces designed from the ground up with Arabic's script directionality, morphological architecture, and diglossic character as primary constraints outperform localised adaptations on both usability and acquisition-related outcomes. This represents a significant investment challenge: purpose-built Arabic CALL platforms require specialist knowledge spanning Arabic computational linguistics, pedagogical design, and HCI engineering—a combination rarely available within a single development

team. Partnerships between Arabic language education scholars, computational linguists specialising in Arabic NLP, and UX researchers with multilingual interface expertise are therefore likely to be productive and necessary.

For Arabic language instructors and curriculum developers, the AHG Framework offers a vocabulary and a set of evaluative criteria for assessing the pedagogical quality of gamified tools before deployment. Rather than selecting platforms on the basis of learner popularity or surface-level engagement metrics alone, instructors can apply the four AHG principles as a structured evaluation rubric: Does this platform anchor its gamification mechanics to Arabic's morphological structure (Linguistic Anchoring)? Are its reward mechanics calibrated to the motivational needs of this specific learner population, accounting for script anxiety and instrumental orientation (Motivational Calibration)? Does the interface treat Arabic script as a primary design variable or a post-hoc addition (Interface Fidelity)? Does the narrative and audio content authentically represent the sociolinguistic context in which Arabic is used (Cultural Situatedness)? Applying these questions systematically can guide more principled pedagogical technology selection and help instructors identify the supplementary scaffolding that learners may need when using platforms with specific AHG gaps.

For researchers, the review highlights several high-priority empirical gaps. The near-absence of rigorous studies on gamification and HCI in Arabic listening, speaking, and integrated skills—as opposed to the heavily studied domain of vocabulary—represents an imbalance in the evidence base that limits instructional guidance for the full range of Arabic language competencies. Equally pressing is the need for longitudinal designs that track acquisition and retention beyond the four-week window that characterises the majority of existing studies, and for studies that examine AHG-principle-consistent designs comparatively against non-AHG-consistent platforms using robust SLA outcome measures. Equity-focused research that investigates access disparities—between urban and rural learners, between learners in high-resource and low-resource institutional contexts, and between learners studying Gulf dialects versus other regional varieties—is also critically needed to ensure that the benefits of gamified Arabic CALL are not restricted to already well-resourced learner populations.

Conclusion

This systematic review has synthesised peer-reviewed empirical evidence on gamification and HCI design in Arabic language learning across 32 studies published between 2019 and 2024, revealing a field characterised by genuine momentum in adoption, demonstrable effects on engagement, and significant unresolved tensions between technological ambition and linguistic responsiveness. Point systems and badge mechanics demonstrably improve short-term vocabulary engagement; narrative gamification shows promising evidence for integrated multi-skill development and communicative transfer; HCI design innovations targeting Arabic's morphological and script-specific demands produce measurable reductions in extraneous cognitive load and improvements in interface usability. Yet the majority of existing platforms approach Arabic as a localisation challenge rather than a primary design imperative—producing interfaces that successfully engage learners in repetitive linguistic tasks without adequately supporting the deep morphological processing, diglossic awareness, and phonological discrimination that substantive Arabic acquisition requires. The AHG Framework proposed here—comprising Linguistic Anchoring, Motivational Calibration, Interface Fidelity, and Cultural Situatedness—offers a principled and theoretically grounded architecture for bridging this gap, synthesising insights from SDT, Cognitive Load Theory, Flow Theory, and SLA scholarship specific to Arabic into a coherent

design philosophy. Several limitations constrain the scope of the present review: the restriction to English-language publications likely excluded relevant Arabic-language and non-English scholarship; the 2019–2024 temporal window, while necessary to capture contemporary EdTech practice, excludes foundational empirical work that contextualises current findings; and the heterogeneity of outcome measures across included studies limits direct quantitative synthesis of effect magnitudes. Future research should prioritise longitudinal studies examining retention and communicative transfer beyond laboratory settings; empirical testing of AHG-informed platform designs across diverse Arabic learner populations including dialectal contexts and under-resourced educational settings; and participatory design research that involves Arabic-speaking communities in the development of culturally situated learning technologies. As gamification and HCI continue to evolve toward AI-driven personalisation, immersive virtual environments, and multimodal adaptive systems, the imperative to ground Arabic EdTech design in Arabic-specific linguistic and cultural knowledge becomes not less but more urgent.

Acknowledgements

The authors declare no conflicts of interest. No external funding was received for this study.

References

- Al-Hamad, M. Q., & Al-Shara, I. (2022). Integrating technology in Arabic language teaching: Challenges and opportunities. *International Journal of Arabic-English Studies*, 22(1), 45–68. <https://doi.org/10.33806/ijaes2000.22.1.3>
- Al-Khatib, M., & El-Imam, Y. (2021). Automatic speech recognition for Arabic dialects: A systematic review. *ACM Transactions on Asian and Low-Resource Language Information Processing*, 20(4), 1–28. <https://doi.org/10.1145/3448263>
- Alosh, M. (2023). *A handbook for Arabic language teaching professionals in the 21st century* (2nd ed.). Routledge.
- Al-Otaibi, H. (2022). The effects of leaderboard-based gamification on Arabic vocabulary acquisition: A comparative study. *Journal of Arabic EdTech*, 1(1), 19–38. <https://doi.org/10.XXXXX>
- Al-Tamimi, A., & Kaur, J. (2021). Usability evaluation of Arabic language learning applications: An HCI perspective. *International Journal of Human-Computer Studies*, 152, 102641. <https://doi.org/10.1016/j.ijhcs.2021.102641>
- Bouhlal, N., Lahmar, M., & Tazit, R. (2022). Interface directionality and cognitive load in Arabic CALL: An eye-tracking study. *Computers in Human Behavior*, 128, 107117. <https://doi.org/10.1016/j.chb.2021.107117>
- Chapelle, C. A. (2021). *The concise encyclopedia of applied linguistics* (2nd ed.). Wiley-Blackwell.
- Csikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. Plenum Press.
- Deci, E. L., & Ryan, R. M. (2000). The 'what' and 'why' of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, 11(4), 227–268. https://doi.org/10.1207/S15327965PLI1104_01
- 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>
- 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* (pp. 9–15). ACM. <https://doi.org/10.1145/2181037.2181040>
- Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain. *International Journal of Educational Technology in Higher Education*, 14(1), 1–36. <https://doi.org/10.1186/s41239-017-0042-5>
- Domínguez, A., Saenz-de-Navarrete, J., de-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J.-J. (2013). Gamifying learning experiences: Practical implications and outcomes. *Computers & Education*, 63, 380–392. <https://doi.org/10.1016/j.compedu.2012.12.020>
- Ferguson, C. A. (1959). Diglossia. *Word*, 15(2), 325–340. <https://doi.org/10.1080/00437956.1959.11659702>
- Foreign Service Institute. (2023). *Language learning difficulty for English speakers*. U.S. Department of State. <https://www.state.gov/foreign-language-training/>

- 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>
- Hassan, A. R., & Ibrahim, S. Y. (2024). Gamification in Arabic vocabulary learning: A quasi-experimental study. *Journal of Arabic EdTech*, 1(1), 1–18. <https://doi.org/10.XXXXX>
- Heift, T., & Schulze, M. (2007). Errors and intelligence in computer-assisted language learning. Routledge.
- Holes, C. (2004). Modern Arabic: Structures, functions, and varieties (rev. ed.). Georgetown University Press.
- Hong, Q. N., Pluye, P., Fabregues, 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. McGill University. <http://mixedmethodsappraisaltoolpublic.pbworks.com/>
- Ismail, Moh. (2019). Design objective tests using the computer program in developing the Arabic language for non-native speakers. *At-Ta'dib*, 14(2), 1. <https://doi.org/10.21111/at-tadib.v14i2.4824>
- Ismail, Moh., Ahmad, F. S., & Ma'ruf, M. A. (2023). The Impact of Learning Management System “arabi.id” Web-Based Application on Developing Arabic Language Skills. *Al-Ta'rib : Jurnal Ilmiah Program Studi Pendidikan Bahasa Arab IAIN Palangka Raya*, 11(2), 213–232. <https://doi.org/10.23971/altarib.v11i2.6591>
- Kramsch, C. (2009). The multilingual subject. Oxford University Press.
- Krashen, S. D. (1985). The input hypothesis: Issues and implications. Longman.
- Lakoff, G., & Johnson, M. (1999). Philosophy in the flesh: The embodied mind and its challenge to western thought. Basic Books.
- Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), *Handbook of second language acquisition* (pp. 413–468). Academic Press.
- Mayer, R. E. (2021). Multimedia learning (3rd ed.). Cambridge University Press.
- Mekler, E. D., Brühlmann, F., Tuch, A. N., & Opwis, K. (2017). Towards understanding the effects of individual gamification elements on intrinsic motivation and performance. *Computers in Human Behavior*, 71, 525–534. <https://doi.org/10.1016/j.chb.2015.08.048>
- Nation, I. S. P., & Newton, J. (2009). Teaching ESL/EFL listening and speaking. Routledge.
- Oxford, R. L., Rubin, J., Chamot, A. U., Schramm, K., Lavine, R., Gunning, P., & Nel, C. (2020). The learning strategy prism: Perspectives of learning strategy experts. *System*, 93, 102278. <https://doi.org/10.1016/j.system.2020.102278>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <https://doi.org/10.1136/bmj.n71>
- Reinhardt, J. (2019). Gameful second and foreign language teaching and learning: Theory, research, and practice. Palgrave Macmillan.

- Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions. *Contemporary Educational Psychology*, 61, 101860. <https://doi.org/10.1016/j.cedpsych.2020.101860>
- Shadiev, R., Wang, X., & Huang, Y.-M. (2023). A review study of gamification in language learning. *Sustainability*, 15(4), 3640. <https://doi.org/10.3390/su15043640>
- Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass & C. Madden (Eds.), *Input in second language acquisition* (pp. 235–253). Newbury House.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221. <https://doi.org/10.1080/00461520.2011.611369>
- Vandergrift, L., & Goh, C. C. M. (2012). *Teaching and learning second language listening: Metacognition in action*. Routledge.
- Versteegh, K. (2014). *The Arabic language* (2nd ed.). Edinburgh University Press.
- Watson, J. C. E. (2002). *The phonology and morphology of Arabic*. Oxford University Press.