The Impact of Gamified Formative Assessment on Writing Complexity, Accuracy, and Fluency: A Macro-Level Investigation in the Iranian EFL Context

Kimia Majidi, Mohammad Javad Rezai [*] [ID], Ali Mohammad Fazilatfar, Seyyedeh Elham Elhambakhsh

Department of English Language and Literature, Faculty of Languages and Literature, Yazd University, Yazd, Iran

Abstract: Coalescing digital technologies into education has foregrounded gamified formative assessment (GFA) as a promising modality for advancing second language acquisition (SLA). The current study explored the effects of GFA on the writing proficiency of Iranian EFL learners, focusing on complexity, accuracy, and fluency (CAF). Employing a quasi-experimental design, intermediate and advanced learners were assigned to either gamified or traditional paper-based formative assessment conditions. Writing CAF performance was evaluated by estimating the mean of macro-constructs of complexity, accuracy, and fluency. Writing performance was analyzed using established CAF measures for complexity, accuracy, and fluency, based on widely recognized frameworks. To assess group and proficiency-level effects, key statistical tests, i.e., Friedman, mixed-design ANOVA, Kruskal-Wallis, Wilcoxon Signed-Rank, independent samples t-test, and Mann-Whitney *U*, were applied. The findings disclosed that GFA enhances writing proficiency, with the most substantial gains observed in complexity and fluency. Accuracy improvements, though evident, were more modest and varied by proficiency level: advanced learners improved more in complexity and fluency, whereas intermediate learners showed greater gains in accuracy. These results accentuate GFA's pedagogical efficacy in fostering engagement, reducing cognitive load, and promoting self-regulated learning in writing instruction, offering crucial insights for language educators, curriculum designers, and digital assessment developers who aim to incorporate innovative, learner-centered assessment strategies.

Keywords: CAF, Composition, Interactive Evaluation, Game-based.

_____

[*] **Corresponding Author.**

Authors' Email Address:

[1] **Kimia Majidi** (kimmajidii@gmail.com), [2] **Mohammad Javad Rezai** (mrezai@yazd.ac.ir), [3] **Ali Mohammad Fazilatfar** (afazilatfar@yazd.ac.ir)), [2] **Seyyedeh Elham Elhambakhsh** (s.e.elhambakhsh@yazd.ac.ir)
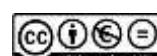
Introduction

The rapid advancement of technology has profoundly transformed pedagogical methods, with gamified learning emerging as a pivotal innovation. Within second language acquisition (SLA) and formative assessment frameworks, gamification has increasingly demonstrated its capacity to enhance motivation, participation, and academic achievement (Deterding et al., 2011; Surendeleg et al., 2014).

Formative assessment, fundamental to monitoring learner progress and fostering self-regulation, has evolved toward more interactive, learner-centered paradigms (Hamari et al., 2014; Wouters et al., 2013). By integrating assessment of learning (AoL) with assessment as learning (AaL), it promotes sustained feedback and cultivates student autonomy through self-reflection (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006), a principle crucial to English as a Second Language (ESL) and foreign language learning (Guskey, 2003; Hamari et al., 2014).

Writing, long regarded as one of the most formidable skills for English as a Foreign Language (EFL) learners, entails considerable cognitive and linguistic complexity (Flower & Hayes, 1981; Hyland, 2003; Richards & Renandya, 2002). Nevertheless, emerging research attests to the potential of gamified formative assessments to alleviate these challenges by providing immediate feedback, reducing test anxiety, and promoting active participation (McLaughlin & Yan, 2017; Surendeleg et al., 2014). Through mechanisms such as rewards, challenges, and progress tracking, gamified assessments foster dynamic, low-stakes environments conducive to iterative learning (Deterding et al., 2011; Dichev & Dicheva, 2017; Li et al., 2023).

In L2 writing assessment, the constructs of complexity, accuracy, and fluency (CAF) are central indicators of linguistic development. Complexity denotes the sophistication and diversity of structures; accuracy, the degree of error-free output; and fluency, the smoothness and efficiency of language production (Ellis, 2008; Skehan, 2009). Recent research inquiries (Bulté & Housen, 2014, 2018; Larsen-Freeman, 2021) have further refined these constructs, incorporating dynamic systems perspectives and multidimensional approaches to capture the evolving nature of L2 proficiency, specifically in response to task-based and technology-mediated instruction. Although gamification has been extensively explored across educational contexts, its specific role in enhancing L2 writing proficiency via formative assessment, particularly in the Iranian EFL context, remains insufficiently addressed. Furthermore, existing research has largely privileged micro-level analyses of CAF

components, often overlooking macro-level performance (Bulté & Housen, 2014; Larsen-Freeman, 2021; Ortega, 2003).

Addressing these gaps, the present study investigates the effects of gamified formative assessment on macro-level CAF measures across intermediate and advanced learners. In doing so, it seeks to enrich the literature on gamification in SLA while offering practical insights for educators and curriculum designers aspiring to implement innovative, learner-centered assessment strategies that enhance writing performance. This research thus lays the groundwork for broader explorations of gamification's potential within EFL and SLA contexts.

Literature Review

Building on the seminal work of Deterding et al. (2011), gamification in education integrates game design elements into non-game contexts to enhance engagement and motivation. Hamari et al. (2014) highlighted its potential to boost intrinsic motivation and learning outcomes, while Wouters et al. (2013) emphasized its alignment with cognitive and emotional learning processes, fostering self-regulation. In language learning, gamification transforms traditional instruction, creating dynamic and interactive experiences.

Writing, a cognitively demanding skill, requires balancing structural complexity, grammatical accuracy, and fluency (Bitchener & Ferris, 2012; Richards & Renandya, 2002). The Complexity, Accuracy, and Fluency (CAF) framework (Ellis, 2008; Housen & Kuiken, 2009; Skehan, 2009) assesses writing performance, with complexity measured by syntactic and lexical sophistication, accuracy by error-free language use, and fluency by output speed and coherence. Housen et al. (2012) confirmed its reliability across second-language contexts, with recent studies (Liu et al., 2024; Bulté & Housen, 2018) advocating for macro-level CAF approaches to comprehensively capture overall performance.

Formative assessment in second language acquisition (SLA) enhances learner autonomy, motivation, and proficiency via continuous feedback (Black & Wiliam, 1998; Nicol & Macfarlane-Dick, 2006). Studies show iterative feedback facilitates skill improvement (Guskey, 2003; Hamari et al., 2014). In writing, formative assessment addresses knowledge gaps, promoting more complex and accurate output (Richards & Renandya, 2002). Gamified formative assessment (GFA) platforms like Quizizz increase engagement, motivating learners and augmenting writing across CAF dimensions. Liu et al. (2024) found that immediate feedback on such platforms enhances self-monitoring, while Zhang and Crawford (2024) highlighted increased fluency through interactive features.

GFA's motivational benefits are crucial for writing improvement. Liu et al. (2024) demonstrated that gamification enhances foreign language enjoyment and the ideal L2 self, increasing learner investment in writing tasks. Competitive and interactive elements extensively maintain interest and promote better writing outcomes (Guo et al., 2024). Immediate feedback facilitates self-regulation and iterative learning, which are deemed essential for writing development (Liu et al., 2024). Gamified approaches in SLA, especially writing, show promise in fostering learner autonomy and cognitive engagement, which is regarded as pivotal for writing development (Dehghanzadeh et al., 2019; Lampropoulos & Kinshuk, 2024; Sailer et al., 2017; Zhang & Hasim, 2023; Zhou & Yu, 2022). Zhang and Crawford (2024) found that gamified assessments improved writing complexity and accuracy by promoting active participation, while game elements like progress tracking reduce cognitive load, enabling focus on complexity and accuracy (Hamari et al., 2014; Surendeleg et al., 2014). Gündüz and Akkoyunlu (2020) noted that gamified assessments foster risk-taking, improving writing fluency, while iterative feedback enhances accuracy (Afifah & Priyana, 2023; Chu & Fowler, 2020; Fan, 2023; Roodi & Slavkov, 2022; Zhang & Hasim, 2023). In contrast, traditional formative assessments may better address grammatical accuracy (Rahimi & Fathi, 2024).

In the Iranian EFL context, research on GFA in writing is limited. Alizadeh & Cowie (2022) showed its effectiveness in improving speaking proficiency and called for more exploration in writing, while Salehi et al. (2023) and Wei et al. (2023) reported that gamification reduced test anxiety and improved writing accuracy through immediate feedback and interactivity. Proficiency level significantly impacts GFA's effectiveness across CAF dimensions (Lambert & Kormos, 2014; Vercellotti, 2017). Gamification's dynamic nature offers differentiated effects based on learners' proficiency, presenting both opportunities and challenges.

Despite increasing research on GFA, few studies have explored its impact on macro-level writing CAF measures in the Iranian context, and little attention has been drawn to how proficiency levels moderate its effects. This study adopts a macro-level approach to examine the impacts of gamified formative assessment on writing performance among intermediate and advanced Iranian EFL learners, focusing on overall performance across complexity, accuracy, and fluency.

The primary research question is: How does gamified formative assessment influence Iranian EFL learners' writing performance, measured by complexity, accuracy, and fluency (CAF), compared to conventional paper-based formative assessment?

Two subordinate questions are addressed:

1. How does GFA affect writing CAF performance in intermediate-level EFL learners compared to traditional assessments?

2. How does GFA affect writing CAF performance in advanced-level EFL learners compared to traditional assessments?

Methods

Research Design

This quasi-experimental study employed a non-equivalent control group design with a pretest-posttest structure to compare the effects of gamified versus traditional paper-based formative assessment on Iranian EFL learners' writing complexity, accuracy, and fluency (CAF). Due to logistical constraints, participants were semi-randomly assigned to experimental and control groups, with English proficiency equivalence ensured via the Quick Oxford Placement Test (QOPT). The study investigated macro-level CAF features across CEFR levels (B1–C2), incorporating within- and between-group comparisons to detect differential impacts by proficiency level.

Participants

Seventy-eight Iranian EFL learners (47 male, 31 female; M_age = 20.87) enrolled in private language institutes in Bandar Abbas participated. Participants were categorized into two proficiency levels of intermediate (B1-B2; $n= 41$) and advanced (C1-C2; $n= 37$) based on their Quick Oxford Placement Test (QOPT) scores. Each proficiency level was further divided into experimental and control groups using a semi-random allocation process to maintain balance, as follows:

- Intermediate: 20 experimental, 21 control

- Advanced: 19 experimental, 18 control

- Overall: 39 experimental, 39 control

Purposive sampling ensured homogeneity in the first language (Persian) and prior English learning experience. Standardized testing (QOPT) confirmed proficiency levels. All participants were non-immersed learners with limited English exposure outside formal education, minimizing confounding variables. Ethical approval was secured, informed consent obtained, and participant anonymity preserved through pseudonymization.

Instruments

Instructional Materials

The *English for Everyone* series, supplemented with TOEFL preparation materials, was deployed as the primary instructional resource. Writing tasks adhered to TOEFL iBT standards to ensure task uniformity.

Gamified Formative Assessment Platforms

The experimental group engaged with Kahoot, Quizizz, Blooket, and Google Forms, selected to compensate for platform-specific limitations and to maximize interactivity, feedback quality, and engagement through gamified elements such as leaderboards, progress tracking, and interactive writing challenges.

- *Kahoot:* Competitive, timed quizzes with real-time scoring.
- *Quizizz:* Synchronous/asynchronous quizzes supporting self-paced learning.
- *Blooket:* Diverse game modes enhancing motivation.
- *Google Forms:* Structured written responses and automated feedback collection.
- Writing tasks were scaffolded to ensure progressive development aligned with pedagogical goals.

Traditional Paper-Based Formative Assessment

The control group completed writing tasks under teacher supervision, received written corrective feedback, and participated in self-review activities. An identical assessment rubric ensured comparability between groups.

Data Collection Procedures

The sixteen-week intervention included two weekly instructional sessions, each lasting approximately 120 minutes. Participants, once per week, were involved in TOEFL-aligned writing tasks inspired by the exercises extracted from the English for Everyone instructional book series, ensuring tasks' authenticity, standardization, and relevance for both groups. Both the experimental and control groups completed parallel forms of the writing assessments, matched for content and difficulty, with the key difference being that the experimental group was assessed weekly through gamified platforms while the control group's evaluation adhered to traditional paper-based assessment deprived of any gamified features.

Pretest and posttest writing assessments, administered under identical conditions, enabled a direct comparison of CAF development. To elaborate, the pre- and post-test tasks

utilized were the authentic TOEFL iBT writing section's tasks. These tests were administered according to the instructions and test procedures established by the Educational Testing Service (ETS) for the TOEFL iBT. Participants received detailed instructions on the assessment procedures before each test. They were permitted to take handwritten notes during the assessment but were not allowed access to the internet, dictionaries, or any other external resources in either group. Written responses were analyzed for complexity (clauses per T-unit, dependent clauses per total clauses), accuracy (error-free T-units and clauses), and fluency (words per T-unit, total words, T-units, and clauses per text).

Feasibility tests were conducted prior to the intervention, identifying limitations in Kahoot's free version and ensuring reliable platform use. Specifically, Kahoot's free version suffers limitations, including the inability to use audio input and restrictions to multiple-choice and true/false question types. To overcome these constraints, an integrated approach employing Kahoot!, Quizizz, Blooket, and Google Forms was adopted to diversify task types and maintain participant engagement. Additionally, engagement of the participants in the experimental group was continuously monitored through online activity logs automatically tracked by the digital platforms deployed. For the control group, the instructor systematically observed and recorded participants' active involvement during sessions to monitor engagement levels.

AI-assisted tools (e.g., Grammarly) were piloted as supplementary supports. This tool was utilized by the course instructor to provide comprehensive feedback to learners in both groups. Participants themselves did not interact directly with Grammarly, as the instructor used it to enhance the quality and consistency of feedback provided. Moreover, inter-rater reliability assessments yielded high agreement ($\alpha > .95$), ensuring coding consistency.

Data Analysis

Writing samples were manually coded by three trained raters based on widely established CAF metrics (Larsen-Freeman, 2006; Norris & Ortega, 2009; Storch & Wigglesworth, 2007; Tavakoli & Skehan, 2005; Wolfe-Quintero et al., 1998), operationalized as follows:

- *Complexity:* Clauses per T-unit; dependent clauses per total clauses.
- *Accuracy:* Error-free T-units per total T-units; error-free clauses per total clauses.
- *Fluency:* Words per T-unit; total words, T-units, and clauses per sample.

A subset (10%) of samples underwent inter-rater reliability checks using Intraclass Correlation Coefficients (ICCs), with calibration sessions enhancing coding precision. Statistical analyses were conducted via SPSS 29.0.2.0. Given violations of normality

assumptions, non-parametric tests (Friedman, Wilcoxon Signed-Rank, Mann-Whitney U) were primarily employed. Kruskal-Wallis tests, with Dunn's post hoc and Bonferroni corrections, assessed multi-group differences. Mixed-design ANOVA evaluated time × group interactions for writing accuracy. Effect sizes were calculated to assess the magnitude of observed effects. A detailed summary is provided in Table 1.

Table 1. Data Analysis Synthesis

Analysis Focus	CAF Dimension	Test(s) Used
Group Differences (Control vs. Experimental)	Macro-Level	Friedman Test, Wilcoxon Signed-Rank Test
	Complementary	
	Complexity	Independent Samples t-Test
	Accuracy	Mann-Whitney U Test
	Fluency	Independent Samples t-Test
CEFR Differences (Intermediate vs. Advanced)	Macro-Level	Kruskal-Wallis Test
	Complementary	
	Complexity	Friedman Test, Wilcoxon Signed-Rank Test, t-Test
	Accuracy	Same as above
	Fluency	Same as above
	Extended Analyses	
	Accuracy (Intermediate)	Mixed-Design ANOVA
	Complexity (Advanced)	Mixed-Design ANOVA

Results

To thoroughly assess the impact of gamified formative assessment on the complexity, accuracy, and fluency (CAF) of Iranian EFL learners' writing performance, a comprehensive statistical analysis was conducted. Descriptive statistics, including means and standard deviations, were calculated for each CAF measure across both the gamified and paper-based assessment conditions at the intermediate and advanced proficiency levels. Means and standard deviations for CAF scores by group (Experimental vs. Control) and test phase (Pre-test vs. Post-test) are as follows:

- Complexity: the experimental group (n = 39) showed a mean increase from 0.91 (SD = 0.13) at the pre-test to 0.96 (SD = 0.09) at the post-test, while the control group's (n = 39) means were 0.95 (SD = 0.11) at the pre-test and 0.94 (SD = 0.13) at the post-test, respectively.

- Accuracy: the experimental group improved from a mean of 0.68 (SD = 0.17) at the pre-test to 0.86 (SD = 0.11) at the post-test, whereas the control group changed from 0.71 (SD = 0.16) to 0.77 (SD = 0.19).

- Fluency: the experimental group's mean increased from 39.06 (SD = 11.32) to 42.65 (SD = 8.92), compared to the control group's mean scores of 40.78 (SD = 10.08) at pre-test and 42.82 (SD = 9.75) post-test.

Moreover, at the CEFR division, changes from the pretest to the post-test favored the experimental group as follows:

- Intermediate level: the experimental group's (n = 20) complexity improved from a mean of 0.87 (SD = 0.15) to 0.93 (SD = 0.09), while the control group (n = 21) remained stable at 0.89 (SD = 0.11) pre- and post-test. As for accuracy, the experimental group displayed improvements from pre- to post-tests (M_{pre} = 0.60, SD_{pre} = 0.19; M_{post} = 0.80, SD_{post} = 0.12); however, for the control group, the participants exhibited relatively stable performance on the pre- and post-tests (M_{pre} = 0.65, SD_{pre} = 0.18; M_{post} = 0.68, SD_{post} = 0.23). Concerning fluency, the experimental group's participants exhibited notable improvements from pre to post-test (M_{pre} = 34.97, SD_{pre} = 8.61; M_{post} = 40.48, SD_{post} = 9.86) while the control group's boost was less considerable (M_{pre} = 37.03, SD_{pre} = 11.05; M_{post} = 38.72, SD_{post} = 9.46).

- Advanced level: complexity means were 0.95 (SD = 0.08) and 0.99 (SD = 0.09) for the experimental group's (n = 19) pre- and post-test, respectively, with control group (n =18) means of 1.01 (SD = 0.06) and 0.99 (SD = 0.09) showing a slight degradation in performance. Additionally, the experimental group's writing accuracy boosted more from pre to post-test (M_{pre} = 0.76, SD_{pre} = 0.09; M_{post} = 0.93, SD_{post} = 0.06) compared to that of the control group (M_{pre} = 0.79, SD_{pre} = 0.08; M_{post} = 0.87, SD_{post} = 0.06). Regarding compositions' fluency, the overall enhancement in performance of the experimental (M_{pre} = 43.36, SD_{pre} = 12.42; M_{post} = 44.93, SD_{post} = 7.39) and the control group (M_{pre} = 45.16, SD_{pre} = 6.75; M_{post} = 47.61, SD_{post} = 7.88) assimilated; nonetheless, the experimental group's performance displayed reduced variability, as traced by the decreased standard deviation.

Data Analysis

In the present study, the assumptions of normality were violated at various stages of both the pre-test and post-test assessments. As a result, non-parametric statistical methods, including the Friedman and Wilcoxon Signed-Rank tests, were employed to analyze the data at the macro level. Additionally, a combination of parametric and non-parametric tests, specifically the independent samples t-test and the Mann-Whitney U test, was used to examine the gain scores between the pre-test and post-test assessments within both the control and experimental groups.

To assess changes in complexity, accuracy, and fluency (CAF) across pre- and post-tests for both experimental and control groups, regardless of CEFR level, Friedman tests were conducted. The results revealed statistically significant differences in CAF scores for both groups: Experimental ($N = 39$, $\chi^2(5) = 161.90$, $p < .000$); Control ($N = 39$, $\chi^2(5) = 170.36$, $p < .000$). Since both groups demonstrated improvements, subsequent analyses were performed using the Wilcoxon Signed-Rank Tests to identify which specific CAF constructs contributed to these changes. The Wilcoxon Signed-Rank Test results, adjusted using the Bonferroni correction ($\alpha = .0167$) to control for the risk of Type I error due to multiple comparisons, showed that all CAF macro constructs significantly improved in the experimental group. In contrast, the control group displayed non-significant and minor improvements in complexity, with accuracy being the only construct showing significant improvement (see Table 2).

Table 2. Friedman Test Inferential Statistics: Group-Level Analysis

Group	χ^2	df	p
Experimental	161.90	5	.000
Control	170.36	5	.000

According to Rosenthal's r, the medium-to-large effect sizes in the experimental group suggest that gamified formative assessment substantially enhanced writing performance, particularly in accuracy. The control group's significant, though smaller, improvement in accuracy indicates a more limited impact of traditional assessment (see Table 3).

AREL

Table 3. Post hoc Wilcoxon Signed Ranks Test: Group-Level Analysis

Posttest-Pretest	*Z-score*	*Asymp. Sig. (2-tailed)*	*r*	
Experimental				
Complexity	-2.92	.004	-.47	Medium
Accuracy	-4.65	.000	-.75	Large
Fluency	-2.74	.006	-.44	Medium
Control				
Complexity	-.858	.391	-.14	Small
Accuracy	-2.46	.014	-.39	Medium
Fluency	-1.51	.132	-.24	Small

Furthermore, between-group comparisons were conducted using Mann-Whitney U and Independent Samples t-Tests. The results of the Mann-Whitney U test for accuracy gains displayed a statistically significant advantage for the experimental group (U = 458.50, Z = -3.02, p = .003, r = -.34), indicating that participants in the experimental group demonstrated significantly greater improvements in accuracy compared to the control group's participants. Independent Samples t-Tests were subsequently deployed to probe variations in complexity and fluency gains between the groups. The analyses confirmed significant differences in complexity, with moderate gains observed in the experimental group (M = 0.05, SD = 0.12), while the control group exhibited minimal change (M = -0.01, SD = 0.11). This statistically significant difference ($t(76)$ = 2.49, p = .015, d = 0.56) reflected a moderate effect size. However, no statistically significant differences between the intervention (M = 3.59, SD = 9.56) and non-intervention (M = 2.04, SD = 8.80) groups were observed concerning fluency gains. ($t(76)$ = 0.75, p = .458, d = 0.17). These results, en masse, accentuate the virtues of gamification in enhancing accuracy and complexity, while fluency failed to display statistically meaningful changes.

To probe the hypothesized differential performance of discrete proficiency levels, a Kruskal-Wallis test was conducted, comparing the following four groups: Experimental-Intermediate (EI), Control-Intermediate (CI), Experimental-Advanced (EA), and Control-Advanced (CA). The analysis identified significant differences in all examined constructs. To elaborate, writing complexity explored through the Kruskal-Wallis test exhibited significant differences in pre-test scores across groups, with $H(3)$ = 23.12, $p < .001$, η^2 = .24. The mean ranks for individual groups were as follows: EI = 25.50, CI = 32.71, EA = 43.50, and CA = 58.75. Both advanced proficiency groups (EA and CA) displayed statistically superior

mean ranks compared to intermediate proficiency groups, underscoring the substantial determining impact of learners' initial proficiency level on writing complexity performance, irrespective of formative assessment type. Notably, the CA group's outperformance hypothesized the dominant effect of prior proficiency rather than formative assessment type on performance at this stage. Analysis of post-test results disclosed remarkably significant differences, $H(3) = 10.43$, $p = .015$, $\eta^2 = .12$, with mean ranks of EI = 35.55, CI = 28.62, EA = 48.03, and CA = 47.58. The experimental-advanced group manifested the most substantial advancement in complexity, followed by the CA group, evidencing both groups' improvements in complexity from pre- to post-test. Additionally, the EI group remarkably excelled in writing complexity compared to the CI group. The considerable development of complexity among EI group members proposed the probable impact of gamified formative assessment on writing complexity; conversely, the absence of marked differences between the advanced proficiency groups suggested that both gamified and paper-based formative assessments were equally effective in promoting the writing complexity of higher proficiency participants.

In a similar vein, the pre-test results analysis of writing accuracy conducted through the Kruskal-Wallis test indicated significant differences across all four groups, $H(3) = 15.62$, $p = .001$, $\eta^2 = .17$, with mean ranks of EI = 27.50, CI = 32.67, EA = 47.68, and CA = 52.17. Due to the outperformance of the advanced groups, particularly the CA group, the association between proficiency level and accuracy can be considered as advanced students have performed better than their intermediate peers in spite of the assessment type. With regards to post-test accuracy scores, more pronounced differences were traced, $H(3) = 31.41$, $p < .001$, $\eta^2 = .31$, with mean ranks of EI = 31.85, CI = 23.10, EA = 61.13, and CA = 44.31. Having the EA group outperforming all others, followed by the CA group and EI group, underscores the positive influence of gamified formative assessment in advancing writing accuracy, especially in the case of advanced learners. Hence, formative assessments that are engaging and interactive, fostering a sustained focus on accuracy, may better benefit higher proficiency learners.

Commencing the analysis of fluency, marked differences were noted in the pre-test, $H(3) = 15.08$, $p = .002$, $\eta^2 = .15$, with mean ranks of EI = 28.80, CI = 32.14, EA = 45.39, and CA = 53.75. Consistent with complexity and accuracy, advanced learners, particularly the CA group, exhibited higher fluency at the pre-test, pointing to a potential association between proficiency level and fluency performance. Fluency assessed on post-test demonstrated significant differences, $H(3) = 11.25$, $p = .010$, $\eta^2 = .10$, with mean ranks of EI = 33.60,

CI = 30.10, EA = 44.34, and CA = 51.92. At this stage, the CA group transcended all others, with EA coming second. Furthermore, EI surpassed the CI group. These results established that, while the control group maintained higher fluency scores, the experimental group, particularly the advanced participants, displayed notable improvements owing to the intervention.

In view of the substantially marked results obtained from Kruskal-Wallis tests, post hoc pairwise comparisons were conducted using Dunn's test with Bonferroni correction ($\alpha < .0083$) to rule out the chances of inflated Type I error resulting from multiple comparisons. The post hoc analyses verified pronounced beneficial effects of gamified formative assessment in improving writing CAF, particularly for advanced learners, with traced substantial improvements in accuracy and fluency of experimental groups. The EA group displayed the most notable improvements with large effect sizes observed in the post-test of accuracy, while the EI participants exhibited marginal improvements. CA maintained higher fluency and pre-test ranks, highlighting the stability of traditional methods in supporting fluency, yet failing to drive statistically significant improvements in complexity and accuracy. The findings propose enhanced cognitive engagement in the experimental group, facilitated through gamification, focusing on the moderating role of proficiency level in determining the success of formative assessment interventions. These findings advocate the pedagogical efficacy of gamified formative assessment in enhancing writing complexity and accuracy, with implications for tailoring interventions according to proficiency levels.

An extended analysis of individual CEFR groups' performance was initially conducted through Friedman and Wilcoxon Signed-Ranks Tests augmented by Mixed-Design ANOVA and Independent Samples *t*-Test run on gain values. The Friedman test conducted on the intermediate-level participants revealed significant performance changes in both the experimental ($N = 20$, $\chi2(5) = 80.83$, $p < .000$) and control groups ($N = 21$, $\chi2(5) = 90.03$, $p < .000$). Post hoc Wilcoxon Signed-Rank tests (see Table 4), with Bonferroni correction ($\alpha = .016$), further demonstrated significant improvements within the experimental group across all three writing CAF dimensions as follows: accuracy ($Z = -2.88$, $p = .004$, $r = -.64$), fluency ($Z = -2.39$, $p = .017$, $r = -.53$), and complexity ($Z = -2.20$, $p = .028$, $r = -.49$). In contrast, the control group exhibited negligible changes in these measures.

Table 4. Wilcoxon Signed-Ranks Test: Intermediate-Level

Posttest-Pretest		*Z-score*	*Asymp. Sig. (2-tailed)*	*r*	
Experimental					
	Complexity	-2.20	.028	-.49	Medium
	Accuracy	-2.88	.004	-.64	Medium
	Fluency	-2.39	.017	-.53	Medium
Control					
	Complexity	-0.26	.794	-.06	Small
	Accuracy	-0.78	.434	-.17	Small
	Fluency	-0.71	.476	-.16	Small

Analysis of intermediate-level participants' writing accuracy was granularly explored through Mixed-Design ANOVA. The results demonstrated a significant main effect of time, $F(1,39) = 9.64$, $p = .004$, $\eta_p^2 = .20$, with accuracy improving from pre-test ($M = 0.63$, $SE = 0.03$) to post-test ($M = 0.74$, $SE = 0.03$). Additionally, a significant time × group interaction was observed, $F(1,39) = 5.07$, $p = .030$, $\eta_p^2 = .12$, indicating superior improvement in accuracy of experimental group (Pre-test: $M = 0.60$, $SE = 0.04$; Post-test: $M = 0.80$, $SE = 0.04$) compared to the control group (Pre-test: $M = 0.65$, $SE = 0.04$; Post-test: $M = 0.68$, $SE = 0.04$). Contrarily, the main effect of the group was found to be non-significant, $F(1,39) = 0.72$, $p = .40$, $\eta_p^2 = .02$. This intimates that while gamification contributed to accuracy gains over time, no notable difference was observed between the groups regarding overall performance. The analysis was advanced by probing gain scores via an independent-samples *t*-test. The analysis of writing complexity gains from pre-test to post-test, although showing a moderate effect size, revealed no significant differences between the experimental group ($M = 0.06$, $SD = 0.14$) and the control group ($M = −0.00$, $SD = 0.12$), $t(39) = 1.65$, $p = .106$, $d = 0.52$. For writing accuracy gains, the results exhibited a moderate effect size but non-significant differences, indicating limited efficacy of gamified formative assessment on the accuracy gains of the control group ($M = 1.69$, $SD = 10.63$) compared to the experimental group ($M = 5.50$, $SD = 10.85$), $t(39) = 1.14$, $p = .262$, $d = 0.36$. Nonetheless, the fluency gains showed significant improvements for the experimental group ($M = 0.20$, $SD = 0.26$) with a large effect size compared to the control group ($M = 0.03$, $SD = 0.21$), $t(39) = 3.91$, $p = .030$, $d = 0.70$. Despite this, following the Bonferroni correction ($\alpha = .016$), the significance of the fluency gains was ruled out, highlighting the importance of cautious interpretation when applying corrections for multiple comparisons. In sum, although the

Friedman and Wilcoxon Signed-Rank Tests confirmed significant developments across all CAF constructs in the intermediate learners' experimental group, particularly concerning accuracy and fluency, the subsequent independent samples *t*-test analyses, after applying the Bonferroni correction, did not confirm the outperformance of the experimental group in terms of overall gain. These findings suggest that gamification, as a formative assessment tool, inherits the potential for advancing writing fluency and accuracy among intermediate learners, while traditional, paper-based methods yielded minimal changes in comparison.

Delving into advanced-level groups' performance, the Friedman test was conducted, revealing substantial differences for both groups: experimental ($N = 19$, $\chi^2(5) = 82.97$, $p < .000$) and control ($N = 18$, $\chi^2(5) = 811.11$, $p < .000$). The subsequent Wilcoxon Signed-Rank post hoc (see Table 5) tests disclosed significant improvements in accuracy ($Z = -3.82$, $p = .000$, $r = -.88$) for the experimental group. Albeit, marginal improvements in writing complexity ($Z = -1.89$, $p = .059$, $r = -.43$) and no significant changes in fluency ($Z = -1.21$, $p = .227$, $r = -.28$) were recorded. As for the control group, there were significant advancements in writing accuracy ($Z = -3.33$, $p = .001$, $r = -.79$) but only negligible changes in complexity ($Z = -1.20$, $p = .231$, $r = -.28$) and fluency ($Z = -1.40$, $p = .145$, $r = -.34$). Given the multiple comparisons, Bonferroni correction ($\alpha = .016$) was applied. After this adjustment, only accuracy remained significantly improved for both groups.

Table 5. Wilcoxon Signed-Ranks Test: Advanced-Level

Posttest-Pretest		*Z-score*	*Asymp. Sig. (2-tailed)*		*r*	
			Experimental			
	Complexity	-2.20	.028		-.49	Medium
	Accuracy	-2.88	.004		-.64	Medium
	Fluency	-2.39	.017		-.53	Medium
Control						
	Complexity	-0.26	.794		-.06	Small
	Accuracy	-0.78	.434		-.17	Small
	Fluency	-0.71	.476		-.16	Small

Analysis of writing complexity among advanced learners was conducted by mixed-design ANOVA to assess the effects of formative assessment type (gamification vs. paper-based) and time (pre-test vs. post-test) on the probed construct. Box's test confirmed the equality of covariance matrices, Box's $M = 1.63$, $p = .676$, and Levene's test indicated

homogeneity of variances ($p > .05$). The main effect of time was not significant, $F(1,35) = 0.43$, $p = .516$, $\eta_p^2 = .01$, $F(1, 35) = 0.43$, $p = .516$, suggesting no overall improvement in complexity over time. However, the interaction between time and group approached significance, $F(1,35) = 3.93$, $p = .055$, $\eta_p^2 = .10$, suggesting that the effect of formative assessment type over time may differ between groups. Pairwise comparisons revealed a significant pre-test difference between groups ($p = .018$), but no significant post-test differences ($p = .974$), indicating that the experimental group showed marginal improvement ($p = .067$), whereas the score of the control group remained stable ($p = .361$). The further prolonged analysis of CAF gain scores of advanced-level participants employed independent-samples t-tests, manifesting remarkable improvements in the complexity of the experimental group ($M = 0.04$, $SD = 0.09$), though this difference was not statistically significant when compared to the control group ($M = 0.02$, $SD = 0.10$), $t(35) = 1.98$, $p = .055$, $d = 0.65$. Regarding fluency, minimal differences were traced between the groups ($t(35) = -0.37$, $p = .713$, $d = 0.12$), with no significant enhancements in either group. However, significant improvements in accuracy were observed for the experimental group ($M = 0.17$, $SD = 0.11$) compared to the control group ($M = 0.08$, $SD = 0.08$), $t(35) = 2.64$, $p = .012$, $d = 0.87$, surpassing the Bonferroni-corrected threshold ($\alpha = .016$).

In summary, the Friedman test revealed significant improvements in writing accuracy among advanced learners in the experimental group, while post hoc Wilcoxon Signed-Rank tests indicated marginal gains in writing complexity. Independent-samples t-tests further confirmed significant accuracy improvements for the advanced experimental group, with fluency remaining stable across both control and experimental groups. Collectively, these findings suggested that gamified formative assessment substantially enhances writing accuracy and complexity at advanced proficiency levels. Although fluency gains were less marked, intermediate learners also benefited, particularly in accuracy. The positive outcomes underscored gamified formative assessment's potential as an effective pedagogical intervention for advancing writing proficiency among EFL learners.

The primary aim of this study was to examine the impact of gamified formative assessment on EFL students' writing performance, focusing on complexity, accuracy, and fluency (CAF), and to assess how this effect varies across proficiency levels. Utilizing a mixed-methods approach, including Kruskal-Wallis, Wilcoxon, and ANOVA analyses, the results highlight the moderating role of proficiency level, with advanced learners enjoying the greatest benefits, while intermediate learners also experience substantial improvements,

particularly in accuracy and fluency. The following summarizes the findings regarding the primary and subordinate research questions.

The Impact of Gamified Formative Assessment on Writing CAF Measures

The study disclosed gamified formative assessment's positive influence on CAF measures for both intermediate and advanced learners. Both groups exhibited considerably profound enhancements in accuracy and complexity under gamified conditions compared to traditional assessments, though fluency gains were less pronounced. Advanced learners demonstrated notable improvements in accuracy and complexity, while intermediate learners showed more substantial advancement in accuracy and complexity, underscoring the enhanced engagement and skill development afforded by gamification. However, fluency remained relatively unchanged, indicating the construct's lower sensitivity to assessment type. The analysis also highlighted proficiency level as a key moderator, with intermediate learners benefiting most from gamified assessments, especially concerning accuracy and complexity. In contrast, advanced learners showed primarily enhanced accuracy, suggesting that their foundational skills in fluency and complexity limit the impact of gamification on these dimensions.

Differential Effects of Gamified Formative Assessment on Writing CAF by Proficiency Level

Intermediate-Level Participants. For intermediate students, gamified formative assessment led to significant gains in accuracy and complexity compared to traditional methods, while fluency improvements were minimal. The experimental group outperformed the control group in accuracy and complexity, emphasizing the effectiveness of gamification for fostering cognitive engagement and skill development at lower proficiency levels.

      Advanced-Level Participants. Among advanced learners, gamified formative assessment notably improved accuracy but had a less pronounced effect on complexity and fluency. The experimental group exhibited substantial gains in accuracy, reinforcing the efficacy of gamification in refining writing accuracy at higher proficiency levels. However, fluency differences between groups were negligible, suggesting that traditional methods may be equally effective at higher proficiency levels. Highlighting the greater refinement of accuracy achieved through gamification, contrarily, the improvements in complexity, though still observable, were less significant.

      The findings underscored the need to tailor assessment methods to learners' proficiency levels. Gamification proved most beneficial for intermediate learners, particularly in

promoting accuracy and complexity, whereas advanced learners gain primarily in accuracy, with minimal changes in fluency and complexity.

In sum, gamified formative assessments can significantly enhance writing proficiency, especially in accuracy and complexity, with varying effects across proficiency levels. The study advocates for integrating gamification into language classrooms, particularly for intermediate learners, to foster greater engagement and improvement. These findings offer valuable insights into the potential of gamification to enhance writing in language education and suggest avenues for future research to optimize its use for fluency and complexity, particularly for advanced learners.

Discussion

The integration of gamified formative assessment (GFA) in English as a Foreign Language (EFL) instruction has garnered significant attention in recent years, with studies exploring its impact on various aspects of language learning. This study contributes to the expanding body of literature on the success of gamification in second language acquisition (SLA) and how proficiency levels mediate its effectiveness in fostering writing proficiency.

At a broader level, this study underscored the transformative potential of gamified formative assessment in promoting writing performance across proficiency levels. The findings suggest that gamified formative assessments can benefit students' writing CAF as a holistic construct, particularly in improving writing complexity and accuracy. However, there are instances where the effects are not uniform across all micro-measures, with some constructs showing statistically significant improvements, while others failing to exhibit substantial changes. This observation aligns with a broader trend in educational research on gamification, where outcomes can be inconsistent depending on the variables assessed and the design of the intervention (Domínguez et al., 2013; Hamari et al., 2014; Sailer et al., 2013; Sailer et al., 2017). Therefore, it is essential to generalize the results vigilantly, especially given the presence of insignificant results in some cases, particularly among intermediate learners.

In support of the current findings, gamification has been shown to have positive effects on student motivation and engagement, which can lead to enhanced learning outcomes (Deterding et al., 2011). Compared to conventional formative assessment, literature confirms that gamified approaches offer a more interactive and dynamic learning environment, fostering increased engagement, motivation, and linguistic output (Deci et al., 2017; Domínguez et al., 2013; Hamari et al., 2014). By integrating game mechanics such as

immediate feedback, competition, and goal-setting, gamified assessment enhances the learning experience and encourages active participation in writing tasks (Koivisto & Hamari, 2019; Seaborn & Fels, 2015; Vlachopoulos & Makri, 2017). For writing assessments, gamification can stimulate cognitive engagement and improve task persistence, among crucial factors required for sustained writing improvement (Caponetto et al., 2014). However, conventional formative assessment, with its emphasis on explicit instruction and controlled practice, may be more successful at fostering grammatical accuracy (Sumida, 2018; Xie & Lei, 2019). Therefore, a blended assessment approach, combining gamified techniques with structured accuracy-focused interventions, may yield optimal results in fostering well-rounded writing proficiency (Golesorkhi & Marandi, 2025; Zhang & Huang, 2024).

The current findings highlight the differential impact of gamified formative assessment on writing CAF performance across proficiency levels. While intermediate learners primarily benefit in writing complexity and fluency, they face challenges in grammatical accuracy, likely due to cognitive overload and the absence of explicit grammar-focused interventions. In contrast, advanced learners exhibit substantial gains in all three dimensions, particularly in accuracy and fluency, as they can better regulate their linguistic output in game-based settings.

To expound further, probing the intermediate learners' (B1, B2) performance, results display moderate improvements in writing complexity and fluency when exposed to gamified formative assessment. Increased complexity, measured by clauses per T-unit and syntactic variety, aligns with existing research suggesting that gamification fosters cognitive engagement and deeper linguistic processing (Hamari et al., 2014; Sailer & Homner, 2020; Yang et al., 2020). The mechanisms underpinning these improvements may include interactive challenges, immediate feedback, and goal-setting, which promote sustained engagement and encourage risk-taking in written production (Dörnyei, 2014, 2020; Sailer et al., 2021; Wang et al., 2024; Wang & Tahir, 2020). This aligns with formative assessment principles that emphasize the role of ongoing feedback and learner autonomy in skill development (Heritage, 2010).

Moreover, a notable enhancement in writing fluency was observed. This result supports the findings by Cheng et al. (2025) and Wang and Li (2025), who argue that gamified elements encourage learners to write more spontaneously by reducing anxiety and fostering motivation. Nonetheless, accuracy gains among intermediate learners were marginal. This finding resonates with previous studies indicating that learners at lower proficiency levels often struggle with grammatical precision in game-based environments due to cognitive

overload (Sweller et al., 2019; Yang, 2024). Unlike traditional formative assessment, which provides structured and explicit grammar instruction, gamified assessments emphasize engagement and active participation, which may lead to trade-offs between fluency and accuracy as supported by Skehan (2009). To address these challenges, Reynolds and Kao (2021) advocate for integrating gamification with explicit corrective feedback, suggesting that real-time error correction can improve grammatical accuracy without sacrificing engagement. Adaptive feedback mechanisms tailored to individual learner needs could serve as an effective pedagogical approach to enhance accuracy among intermediate EFL students. Although efforts have been made to provide participants with real-time corrective feedback, this aspect of gamification should be further augmented. Additionally, a hybrid model that incorporates gamified exercises alongside explicitly structured grammar instruction may optimize accuracy gains while maintaining the motivational benefits of gamification (Mohamed et al., 2024; Hong et al., 2022). Future research should explore the sustainability of these interventions over extended learning periods to assess their long-term impact.

Scrutinizing the performance of advanced learners (C1, C2), significant strides were charted across all CAF dimensions when engaged in gamified formative assessment. Unlike intermediate learners, accuracy gains were pronounced for advanced learners, corroborating the assertion that higher-proficiency learners are better equipped to self-monitor and refine their linguistic output in interactive digital environments (Godwin-Jones, 2014; Li & Hegelheimer, 2013; Rahimi & Fathi, 2024; Vlachopoulos & Makri, 2017). The presence of immediate feedback mechanisms in gamified assessment likely contributed to the improved grammatical precision and coherence observed in participants' written production (Deterding et al., 2011; Li et al., 2024; Wei et al., 2023). Furthermore, pronounced enhancements in fluency were manifest, with advanced learners generating more elaborate and lexically diverse compositions. This is in concert with studies by Alshuaifan (2024) and Yavuz et al. (2020), proposing that gamification promotes spontaneous language production by lowering the affective filter and creating an immersive learning experience. Moreover, the cognitive engagement required to navigate gamified challenges encourages advanced learners to generate complex ideas more fluidly, thereby improving their overall writing performance (Gee, 2003; Zhang & Hasim, 2023). From a complexity perspective, advanced learners demonstrated higher levels of syntactic depth and lexical diversity, substantiating the findings of Sailer and Homner (2020), underscoring gamification's potential to foster sophisticated linguistic expression. However, task complexity must be carefully calibrated to ensure that the cognitive demands of gamified tasks do not outweigh their benefits (Zou et al., 2024).

AREL

The integration of scaffolding techniques, such as adaptive difficulty levels and personalized challenge settings, may optimize learning outcomes.

Conclusions

In sum, this study underscores the importance of proficiency level in mediating the effectiveness of gamified formative assessment. For intermediate learners, while gains in complexity and fluency are evident, accuracy remains a challenge, which calls for more targeted feedback mechanisms and explicit grammar instruction. On the other hand, advanced learners benefit from gamified formative assessment across all three CAF dimensions, especially in terms of accuracy and fluency, where their higher proficiency allows for better regulation of their linguistic output. Given these findings, the integration of gamification into EFL classrooms appears to hold significant promise, but a hybrid model that combines gamified techniques with explicit instruction may yield the most comprehensive improvements in writing performance.

Furthermore, this study aligns with formative assessment theories (Black & Wiliam, 1998) and gamification principles grounded in self-determination theory (Deci & Ryan, 2000), contributing to a nuanced understanding of their intersection in EFL instruction.

Limitations

Despite its methodological rigor, the current study has several limitations. First, while the quasi-experimental design is robust, it cannot fully eliminate potential extraneous variables, such as individual differences in digital literacy. Second, the intervention period was limited by institutional scheduling, which may not have allowed for the capture of long-term retention effects. Lastly, although CAF metrics offer a valuable macro-analytic perspective, they do not account for the micro-level cognitive processes involved in writing development. Notwithstanding these limitations, this study provides important empirical insights into the role of gamified formative assessment in EFL writing development. The findings have significant pedagogical implications for curriculum designers and educators seeking to integrate gamified strategies into formative assessment practices.

Implications

Ultimately, implementing gamified formative assessments (GFA) in educational settings offers numerous advantages, namely positive learning experiences (Zhang & Crawford, 2024) along with boosted motivation, engagement, critical thinking, and self-monitoring

(Liu et al., 2024). Nevertheless, several challenges merit consideration. Technical issues, for instance, connectivity problems and platform accessibility, can impede the seamless integration of GFA into curricula. Additionally, the competitive nature of gamified assessments may induce anxiety among some learners, potentially adversely affecting their performance. It is also essential to ensure that game elements do not overshadow educational objectives, maintaining a balance between engagement and learning outcomes (Domínguez et al., 2013). Furthermore, the long-term effectiveness of gamified assessment remains an open question. Studies suggest that while gamification enhances short-term motivation, its sustained impact on writing performance requires further investigation (Deterding et al., 2011; Dichev & Dicheva, 2017; Li et al., 2024). Future research should explore the longitudinal effects of gamified writing assessment, examining how learners' motivation, cognitive engagement, and writing proficiency evolve over time in response to game-based interventions. Besides, research focus can be directed toward studies that delve into how gamification can be systematically integrated into existing curricula, ensuring that educators receive adequate training in implementing effective gamified interventions. Moreover, exploring hybrid models that integrate gamification with traditional assessment approaches may offer a more comprehensive strategy for fostering writing proficiency among EFL learners.

Acknowledgments

Declaration of Conflicting Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Afifah, A., & Priyana, J. (2023). How can gamification help language learners? Insight from EFL students and teachers. *International Journal of Educational Research & Social Sciences, 5*(4), 867–876. https://doi.org/10.51601/ijersc.v5i4.867

Alizadeh, M., & Cowie, N. (2022). Language learning and virtual reality: A scoping review. In S. Wilson, N. Arthars, D. Wardak, P. Yeoman, E. Kalman, & D.Y.T. Liu (Eds.),

Reconnecting relationships through technology. *Proceedings of the 39th International Conference on Innovation, Practice and Research in the Use of Educational Technologies in Tertiary Education, ASCILITE 2022 (Article e22258).* https://doi.org/10.14742/apubs.2022.258

Alshuaifan, A. S. (2024). Effectiveness of a gamification-based enrichment program in developing EFL creative writing skills among the intermediate school students. *Journal of Educational and Psychological Researches, 21*(81), 650–702. https://doi.org/10.52839/0111-000-081-022

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–74. https://doi.org/10.1080/0969595980050102

Bitchener, J., & Ferris, D. R. (2012). *Written corrective feedback in second language acquisition and writing.* Routledge. https://B2n.ir/wy5262

Bulté, B., & Housen, A. (2014). Conceptualizing and measuring short-term changes in L2 writing complexity. *Journal of Second Language Writing, 26*, 42–65. https://doi.org/10.1016/j.jslw.2014.09.005

Bulté, B., & Housen, A. (2018). Syntactic complexity in L2 writing: Individual pathways and emerging group trends. *International Journal of Applied Linguistics, 28*(1), 147–164. https://doi.org/10.1111/ijal.12196

Caponetto, I., Earp, J., & Ott, M. (2014, October). Gamification and education: A literature review. In M. Marczewski (Ed.), *Proceedings of the 8th European Conference on Games Based Learning* (pp. 50–57). Academic Conferences International Limited. https://B2n.ir/jy4823

Cheng, J., Lu, C., & Xiao, Q. (2025). Effects of gamification on EFL learning: A quasi-experimental study of reading proficiency and language enjoyment among Chinese undergraduates. *Frontiers in Psychology, 16.* https://doi.org/10.3389/fpsyg.2025.1448916

Chu, M. W., & Fowler, T. A. (2020). Gamification of formative feedback in language arts and mathematics classrooms: Application of the learning error and formative feedback (LEAFF) model. *International Journal of Game-Based Learning, 10*(1), 1–18. https://doi.org/10.4018/IJGBL.2020010101

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

Deci, E. L., Ryan, R. M., & Koestner, R. (2017). Intrinsic motivation in education: A critical review of gamification research. *Educational Psychology Review, 29*(2), 213–240. https://doi.org/10.1007/s10648-017-9405-7

Dehghanzadeh, H., Fardanesh, H., Hatami, J., Talaee, E., & Noroozi, O. (2019). 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. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, (pp. 9–15). https://doi.org/10.1145/2181037.2181040

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A critical review. *International Journal of Educational Technology in Higher Education, 14*(1), 9. 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 results. *Computers and Education, 63*, 380–392. https://doi.org/10.1016/j.compedu.2012.12.020

Dörnyei, Z. (2014). *Motivational strategies in the language classroom*. Cambridge University Press.

Dörnyei, Z. (2020). *Innovations and challenges in language learning motivation*. Routledge. https://doi.org/10.4324/9780429485893

Ellis, R. (2008). *The study of second language acquisition* (2nd ed). Oxford University Press. https://B2n.ir/fz2842

Fan, N. (2023). Exploring the effects of automated written corrective feedback on EFL students' writing quality: A mixed-methods study. *SAGE Open, 13*(2), 1–15. https://doi.org/10.1177/21582440231181296

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication*, *32*(4), 365–387. https://doi.org/10.2307/356600

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. Palgrave Macmillan. https://B2n.ir/be5833

Godwin-Jones, R. (2014). Games in language learning: Opportunities and challenges. *Language Learning & Technology, 18*(2), 9–19. https://B2n.ir/sp7117

Golesorkhi, Z., & Marandi, S. S. (2025). Exploring the impact of gamified web-based collaborative writing on individual L2 writing performance in higher education. *Computer-Assisted Language Learning Electronic Journal, 26*(1), 1–20. https://doi.org/10.54855/callej.252611

Guo, K., Li, Y., & Chu, S. K. W. (2024). Game on: enhancing argumentative writing with digital gamification. *Innovation in Language Learning and Teaching*, 1-12. https://doi.org/10.1080/17501229.2024.2374347

Gündüz, A. Y., & Akkoyunlu, B. (2020). The effectiveness of gamification in flipped learning: A quasi-experimental study with undergraduate students. *SAGE Open, 10*(3), 1–16. https://doi.org/10.1177/2158244020979837

Guskey, T. R. (2003). How classroom assessments improve learning. *Educational Leadership, 60*(5), 6-11. https://B2n.ir/jg1703

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

Heritage, M. (2010). *Formative assessment and next-generation assessment systems: Are we losing an opportunity?*. Council of Chief State School Officers. https://eric.ed.gov/?id=ED543063

Hong, J. C., Hwang, M. Y., Liu, Y. H., & Tai, K. H. (2022). Effects of gamifying questions on English grammar learning mediated by epistemic curiosity and language anxiety. *Computer Assisted Language Learning, 35*(7), 1458-1482. https://doi.org/10.1080/09588221.2020.1803361

Housen, A., & Kuiken, F. (2009). Complexity, accuracy and fluency in second language acquisition. *Applied Linguistics*, *30*(4), 461-473. https://doi.org/10.1093/applin/amp048

Housen, A., Kuiken, F., & Vedder, I. (Eds.). (2012). *Dimensions of L2 performance and proficiency: Complexity, accuracy and fluency in SLA*. John Benjamins. https://doi.org/10.1075/lllt.32

Hyland, K. (2003). *Second language writing*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667251

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

Lambert, C., & Kormos, J. (2014). Complexity, Accuracy, and Fluency in Task-based L2 Research: Toward More Developmentally Based Measures of Second Language Acquisition. *Applied Linguistics, 35*(5), 607-614. https://doi.org/10.1093/applin/amu047

Lampropoulos, G., & Kinshuk. (2024). Virtual reality and gamification in education: A systematic review. *Educational Technology Research and Development, 72*(4), 1691–1785. https://doi.org/10.1007/s11423-024-10351-3

Larsen-Freeman, D. (2006). The emergence of complexity, fluency, and accuracy in the oral and written production of five Chinese learners of English. *Applied Linguistics, 27*(4), 590-619. https://doi.org/10.1093/applin/aml029

Larsen-Freeman, D. (2021). Complexity and L2 development: A dynamic systems theory approach. *Language Learning, 71*(S1), 1–36. https://doi.org/10.1111/lang.12444

Li, C., Fryer, L. K., & Chu, S. K. W. (2024). Utilising gamified formative assessment to support English language learning in schools: A scoping review. *Innovation in Language Learning and Teaching, 19*(2), 188–204. https://doi.org/10.1080/17501229.2024.2352789

Li, M., Ma, S., & Shi, Y. (2023). Examining the effectiveness of gamification as a tool promoting teaching and learning in educational settings: A meta-analysis. *Frontiers in Psychology, 14*, Article 1253549. https://doi.org/10.3389/fpsyg.2023.1253549

Li, Z., & Hegelheimer, V. (2013). Mobile-assisted grammar exercises: Effects on self-editing in L2 writing. *Language Learning & Technology, 17*(3), 135–156. http://llt.msu.edu/issues/october2013/lihegelheimer.pdf

Liu, G. Z., Fathi, J., & Rahimi, M. (2024). Using digital gamification to improve language achievement, foreign language enjoyment, and ideal L2 self: A case of English as a foreign language learners. *Journal of Computer Assisted Learning, 40*(4), 1347–1364. https://doi.org/10.1111/jcal.12954

McLaughlin, T., & Yan, Z. (2017). Diverse delivery methods and strong psychological benefits: A review of online formative assessment. *Journal of Computer Assisted Learning, 33*(6), 562–574. https://doi.org/10.1111/jcal.12200

Mohamed, K. A., El-Naggar, M. E., & Alazab, G. S. (2024). Effectiveness of gamified cooperation and competition in blended learning environment for EFL business writing in TVET. *European Scientific Journal, ESJ, 20*(17), 107. https://doi.org/10.19044/esj.2024.v20n17p107

Nicol, D., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *126566Studies in Higher Education, 31*(2), 199–218. https://doi.org/10.1080/03075070600572090

Norris, J. M., & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics, 30*(4), 555-578. https://doi.org/10.1093/applin/amp044

Ortega, L. (2003). Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing. *Applied Linguistics, 24*(4), 492–518. https://doi.org/10.1093/applin/24.4.492

Rahimi, M., & Fathi, J. (2024). Exploring the impact of automated written corrective feedback on the academic writing skills of EFL learners: An activity theory perspective. *Education and Information Technologies, 30*(3), 2691–2735. https://doi.org/10.1007/s10639-024-12896-5

Reynolds, B. L., & Kao, C. W. (2021). The effects of digital game-based instruction, teacher instruction, and direct focused written corrective feedback on the grammatical accuracy of English articles. *Computer Assisted Language Learning, 34*(4), 462–482. https://doi.org/10.1080/09588221.2019.1617747

Richards, J. C., & Renandya, W. A. (Eds.). (2002). *Methodology in language teaching: An anthology of current practice*. Cambridge University Press. https://doi.org/10.1017/CBO9780511667190

Roodi, F., & Slavkov, N. (2022). Gamification in L2 teaching and learning: Linguistic risk-taking at play. *OLBI Journal, 12*(1), 1–21. https://doi.org/10.18192/olbij.v12i1.6000

Sailer, M., Hense, J., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture(s) Journal, 19*, 28-37. https://core.ac.uk/download/pdf/26951671.pdf

Sailer, M., Hense, J., Mayr, S. K., & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior, 69*, 371–380. https://doi.org/10.1016/j.chb.2016.12.033

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

Sailer, M., Murböck, J., & Fischer, F. (2021). Digital learning in schools: What does it take beyond digital technology?. *Teaching and Teacher Education, 103*, 103346. https://doi.org/10.1016/j.tate.2021.103346

Salehi, A. M., Mohammadi, H. A., Jenabi, E., Khanlarzadeh, E., & Ashtari, K. (2023). Quality of evidence and pedagogical strategy in using gamification in medical education literature: A systematic review. *Simulation & Gaming, 54*(6), 842–865. https://doi.org/10.1177/10468781231195903

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

Skehan, P. (2009). Modelling second language performance: Integrating complexity, accuracy, fluency, and lexis. *Applied Linguistics*, *30*(4), 510-532. https://doi.org/10.1093/applin/amp047

Storch, N., & Wigglesworth, G. (2007). *Writing tasks: Comparing individual and collaborative writing*. In M. del Pilar Garcia-Mayo (Ed.), *Investigating tasks in formal language learning*. Multilingual Matters. https://doi.org/10.21832/9781853599286-011

Sumida, A. (2018). The effects of explicit grammar instruction and written corrective feedback on accuracy development in EFL writing. *Annual Review of English Language Education in Japan, 29*, 17–32. https://doi.org/10.20581/arele.29.0_17

Surendeleg, G., Murwa, V., Yun, H. K., & Kim, Y. S. (2014). The role of gamification in education – A literature review. *Contemporary Engineering Sciences*, *7*(33), 1609–1616. http://dx.doi.org/10.12988/ces.2014.411217

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Tavakoli, P., & Skehan, P. (2005). Strategic planning, task structure, and performance testing. In R. Ellis (Ed.), *Planning and task performance in a second language* (pp. 239–273). John Benjamins.

Vercellotti, M. L. (2017). The development of complexity, accuracy, and fluency in second language performance: A longitudinal study. *Applied Linguistics*, *38*(1), 90-111. https://doi.org/10.1093/applin/amv002

Vlachopoulos, D., & Makri, A. (2017). The effect of games and simulations on higher education: A systematic literature review. *International Journal of Educational Technology in Higher Education*, *14*(1), 22. https://doi.org/10.1186/s41239-017-0062-1

Wang, S., Kong, X., & Wang, N. (2024). Gamification for learning: Development and application of learning software for enhancing student engagement and motivation. In *Proceedings of the 2024 13th International Conference on Educational and*

Information Technology (ICEIT) (pp. 61–66). IEEE. https://doi.org/10.1109/ICEIT61397.2024.10540849

Wang, C., & Li, Q. (2025). Effects of gamification on Chinese EFL college students' writing error awareness and writing performance. *International Journal of English Linguistics, 15*(1), 76–89. https://doi.org/10.5539/ijel.v15n1p76

Wang, A. I., & Tahir, R. (2020). The effect of using Kahoot! for learning – A literature review. *Computers & Education, 149*, 103818. https://doi.org/10.1016/j.compedu.2020.103818

Wei, P., Wang, X., & Dong, H. (2023). The impact of automated writing evaluation on second language writing skills of Chinese EFL learners: a randomized controlled trial. *Frontiers in Psychology, 14*. https://doi.org/10.3389/fpsyg.2023.1249991

Wolfe-Quintero, K., Inagaki, S., & Kim, H. Y. (1998). *Second Language Development in Writing: Measures of Fluency, Accuracy, and Complexity*. University of Hawai'i, Second Language Teaching and Curriculum Center.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, *105*(2), 249-265. https://doi.org/10.1037/a0031311

Xie, Q., & Lei, Y. (2019). Formative assessment in primary English writing classes: A case study from Hong Kong. *Asian EFL Journal, 23*(5), 55–95. https://B2n.ir/yf4019

Yang, D. (2024). Cognitive load theory: A bibliometric study over a decade and pedagogical implications. *Journal of Education and Educational Research, 10*(1), 357–361. https://doi.org/10.54097/q7eyhd58

Yang, Q., Chang, S., Hwang, G., & Zou, D. (2020). Balancing cognitive complexity and gaming level: Effects of a cognitive complexity-based competition game on EFL students' English vocabulary learning performance, anxiety, and behaviors. *Computers & Education, 148*, 103808. https://doi.org/10.1016/j.compedu.2020.103808

Yavuz, F., Özdemir, E., & Çelik, Ö. (2020). The effect of online gamification on EFL learners' writing anxiety levels: A process-based approach. *World Journal on Educational Technology: Current Issues, 12*(2), 62–70. https://doi.org/10.18844/wjet.v12i2.4600

Zhang, Z., & Crawford, J. (2024). EFL learners' motivation in a gamified formative assessment: The case of Quizizz. *Education and Information Technologies, 29*, 6217-6239. https://doi.org/10.1007/s10639-023-12034-7

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

Zhang, Z., & Huang, X. (2024). *Exploring the impact of the adaptive gamified assessment on learners in blended learning. Education and Information Technologies, 29*, 21869–21889. https://doi.org/10.1007/s10639-024-12708-w

Zou, Y., Kannan, S., & Sidhu, G. K. (2024). Influence of task complexity on text features and writing scores: Evidence from college students in Southern China. *SAGE Open*, *14*(2), 21582440241284186. https://doi.org/10.1177/21582440241284186

Zhou, Z., & Yu, Z. (2022). The impact of gamification on the time-limited writing performance of English majors. *Education Research International, 2022*, Article 4650166. https://doi.org/10.1155/2022/4650166