# Exploring the Impact of Gamified Web-Based Collaborative Writing on Individual L2 Writing Performance in Higher Education

Zahra Golesorkhi[1], S. Susan Marandi[1*]

[1] Department of English, Faculty of Literature, Alzahra University, Tehran, Iran
[*]Corresponding Author's Email: susanmarandi@alzahra.ac.ir
[*] https://orcid.org/0000-0001-9852-1880
https://doi.org/10.54855/callej.252611



Received: 16/02/2024 Revision: 12/09/2024 Accepted: 22/12/2024 Online: 17/05/2025

## ABSTRACT

The present study intends to investigate the effect of using a gamified platform called Classcraft on the individual L2 writing performance of Iranian university students of TEFL. For this purpose, a group of 120 male and female intermediate university students was assigned to three treatment groups as well as a control group, each containing 30 participants, in a quasi-experimental pretest/post-test design. The first experimental group benefited from both gamification and collaboration while interacting via the Classcraft platform to do the tasks. The second experimental group did the writing tasks individually via the same platform, while the third group practiced collaborative writing without gamification. Finally, the control group did the writing assignments without either gamification or collaboration. The results of the two-way ANCOVA revealed that the gamification factor significantly improved the writing performance of the students who utilized the gamified platform (i.e., the first and second experimental groups). However, the collaboration factor failed to exert a statistically significant impact on the scores of the participants who did the tasks jointly (i.e., the first and third experimental groups). The findings of this study can provide practical implications for language teachers, materials designers and policy-makers in education.

**Keywords**: Gamification, Collaborative Writing, L2 Writing Performance, Classcraft

## Introduction

Research indicates that, similar to most other countries, Iranian students of English as a Foreign Language (EFL) often face problems in understanding and applying L2 writing skills (Hashemi et al., 2010; Rezaei & Jafari, 2014). As Memari (2021) observes, Iranian L2 learners typically have problems identifying and consequently producing the most important elements of writings, i.e., thesis statements, topic sentences, supporting sentences, and conclusions. In addition, as Marandi and Seyyedrezaie (2017) note, traditional academic writing classes often turn out to be frustrating in terms of improving the students' writing ability, and thus make their motivation in L2 writing decrease due to the stressful settings and the writing anxiety the students deal with in class. This is while employing new technological tools in teaching and

CITATION | 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 (CALL-EJ),* 26(1), 1-20. DOI: https://doi.org/10.54855/callej.252611

learning L2 writing has introduced new horizons to L2 learners and teachers. Specifically, there is an increasing body of research on the potential effects of incorporating gamification via technology as an effective strategy to uphold the learners' interest and engagement in the learning process and eventually encourage them to move towards self-learning (Anunpattana, 2021; Jeon, 2022; Rincon-Flores et al., 2022). Moreover, gamification is believed to alleviate the students' foreign language writing anxiety to a considerable extent (Yavuz et al., 2020).

In addition to utilizing technological tools in teaching L2 writing, employing collaborative writing has also been highly recommended as an effective teaching method in L2 writing classes (Dobao, 2012; Storch, 2011). The interest in L2 collaborative writing can be witnessed in the numerous studies done, such as studies investigating collaboration between expert and novice students (Lee, 2004), the effect of the environment (Kessler, 2009), task-type (Aydin & Yildiz, 2014), application of online tools (Kessler, 2009; Lee, 2004), and teacher intervention in collaboration (Kessler, 2009). However, implementing collaborative writing in face-to-face L2 classes is often hindered by limitations such as time constraints and the difficulty of monitoring such activities by the teacher during class time (Rezai, 2022), and very few research projects appear to have focused on the impact of gamified web-based collaborative writing. The current study aims to explicitly explore the extent to which this approach can help improve Iranian learners' *individual* L2 writing performance.

## Literature Review

There is a body of accumulated research on L2 writing instruction with the use of technology, indicating that employing digital tools, particularly gamified ones, improves the quality of learning about scientific writing (Kalogiannakis, 2021) and creative writing (Verswijvelen, 2020), as well as academic writing (Tantawi et al., 2018; Zhihao & Zhonggen, 2022). The following sections briefly reflect on the main variables of the current study (i.e., collaborative writing and gamification) to provide a more comprehensive view of the present investigation.

### *The role of collaborative writing in language learning*

Collaborative writing is defined as "an activity where there is a shared and negotiated decision-making process and a shared responsibility for the production of a single text" (Storch, 2013, p. 3). The negotiation and interaction among learners as they focus on producing a written text lay the foundation for collaborative writing (Tanis, 2020). The impact of interaction among peers and its significance in learning has been widely recognized. According to Dao (2020), collaborative learning and interacting with peers can positively affect problem-solving abilities, information-seeking skills, and foster learner engagement.

Theoretically speaking, collaborative learning is substantiated by the notion of scaffolding and the socio-cultural theory (Vygotsky, 1978), which lays emphasis on the role of peers in triggering each other's potential level of development. As Lundstrom and Baker (2009) have pointed out, peer interaction and collaborative writing can effectively enhance the learners' writing competencies. Specifically, when students negotiate and interact with each other during peer editing (i.e., a process of collaborative learning), they become enabled to critique their peers' writings, and in turn, their own written productions (Yu & Lee, 2016).

On the other hand, however, comparative studies on collaborative writing have provided contradictory results regarding whether it has a more positive effect than individual writing. It appears that collaborative writing is better for enhancing students' grammatical accuracy in their writing production, which leads to a better understanding of lexis-related problems (Wigglesworth & Storch, 2009). However, the study by Zabihi et al. (2013) suggested that writing fluency did not show a significant improvement in students working in collaborative

groups compared to those working individually. Moreover, instructors have expressed concerns about the fairness and reliability of assigning group composition and determining whether each student has equally participated in the writing activities (Strauss & U, 2007). Thus, this type of practice has been both recommended and criticized in various studies, and the best approach toward employing this method seems to be the implementation of collaborative writing with an awareness of the potential benefits and downsides, and addressing possible deficiencies through adopting appropriate strategies (Lopres et al., 2023). In this study, collaborative writing was employed as an L2 teaching method to observe the effects it can exert on the students' L2 writing performance alongside gamification.

### *Gamification and language learning*

According to experiential learning theory, gamification can be described as incorporating game elements into a non-game context in order to enhance the learners' motivation and engagement (Deterding et al., 2011). These fundamental elements, present in nearly every game, include a clearly defined objective, a set of rules to achieve this objective, a feedback system that rewards desired performance as players attempt to unlock more achievements, and voluntary involvement, so that players are not obligated to attempt each task (Morris et al., 2013).

Gamification and digital games have been noted by many studies as increasing learning motivation and outcomes (e.g., Wiethof et al., 2021). The educational use of gamification in class has been extensively studied by many researchers, such as Cronk (2012), Deterding (2012), and Stott and Neustaedter (2013), who have attributed a positive role to gamification in education, directly increasing the learners' level of motivation and engagement. As Reynolds & Taylor (2020) have pointed out, technology has enhanced the visual aspect of learning by presenting materials in a more engaging and interactive manner. This, coupled with rapid feedback which has enabled learners to receive immediate corrections to their work, can lead to faster progress and mastery (Zhou et al., 2017). Furthermore, digital devices have made self-study and personalized learning more accessible, enabling learners to take control of their own learning (Schmid et al., 2023). The use of bite-sized lessons has also made learning more manageable (Mazlan et al., 2023).

In the Iranian context, Salimi and Zanganeh (2022) investigated the effect of gamification on English vocabulary learning among fifth-grade students. They implemented gamification in their English course by using an Iranian vocabulary mobile application called *Haft-Khan-e Esfandiar [i.e., Esfandiar's Seven Labors]*. They concluded that gamification improved English vocabulary learning among the participants. In another study, Liu et al. (2024) explored the effects of digital gamified language learning using the same mobile application (i.e., *Haft-Khan-e Esfandiar)* on Iranian EFL learners' language achievement. Their results revealed that the digital learners outperformed their non-digital counterparts in all measures. Also, Vameghshahi and Ghonsooly (2023) designed an educational video game, entitled *Lost p*, which utilized a process-based approach to improve learners' writing ability. The results of their study showed that the experimental group demonstrated a better L2 writing performance compared to the control group. Overall, there is some evidence that gamification can effectively facilitate the teaching and learning of L2 writing in different contexts. In the present study, gamification and collaborative writing are employed together as a teaching method to help students improve their L2 writing performance.

### *Research questions*

This study attempts to address the following research questions:

1. Does using a gamified web-based learning platform (i.e., Classcraft) have a significant effect on the individual L2 writing performance of Iranian university students of TEFL?
2. Does collaborative L2 writing have a significant effect on the individual L2 writing performance of Iranian university students of TEFL?
3. Is there any significant interaction between gamification and collaborative L2 writing?

## Method

This section includes descriptions of participants, data collection and data analysis procedures, instrumentation and measurements, as well as the research design.

### *Participants*

The population addressed in this study was Iranian university students of TEFL. As random sampling was not feasible, the participants of this study were selected using a convenience sampling method. The initial sample included 180 male and female students of a number of English essay writing classes at Farhangian University (i.e., Shahid Bahonar and Martyrs of Mecca Higher Education Center) majoring in TEFL at the B.A. level. The age range of the participants was between 19 to 24 years old, and their English proficiency level, assessed via a DIALANG placement test, was generally at an intermediate level. The score range of this test is presented in Table 1 below.

**Table 1.**

*Score range of the DIALANG test (DIALANG, 2024)*

| Score range | Corresponding CEFR Level | Level Descriptors |
|---|---|---|
| 0 – 100 | A1 | Basic |
| 101 – 200 | A2 | Elementary |
| 201 – 400 | B1 | Low Intermediate |
| 401 – 600 | B2 | High Intermediate |
| 601 – 900 | C1 | Low Advanced |
| 901 – 1000 | C2 | High Advanced |

As the mean score of the participants was 540.73, and considering their score range (426 – 658), it can be concluded that the majority of the students could be placed in the intermediate level. Those who were not, were removed from the study, leaving us with 120 participants.

### *Materials and Instruments*

The following instruments were used in the present study to collect the relevant data.

### *DIALANG Test*

DIALANG is an online diagnostic testing system which is designed to evaluate a person's overall proficiency in 14 European languages. Competences tested are reading, writing, listening, grammar and vocabulary, while speaking is excluded for technical reasons (DIALANG, 2024). The participants in this study took the writing test in order to be homogenized to match the target population. According to the creators of the test, DIALANG has been developed based on established language proficiency frameworks and is both valid and reliable (DIALANG, 2024). In terms of validity, the DIALANG test has been designed to

align with the Common European Framework of Reference for Languages (CEFR), and is designed to be adaptable to different contexts and purposes, which increases its validity (European Commission, 2018). Regarding reliability, the DIALANG test has been subject to extensive piloting and testing, and its reliability has been established through statistical analyses. The test is designed to be self-administered and self-assessed, which may increase the reliability of the results as it minimizes the potential for rater bias (Sari & Erten, 2021).

### *Writing Pretest and Post-test*

A pre- and post-test design was used to measure individual L2 writing performance of the participants. Test-effect was minimized by assigning students different writing prompts for pre- and post-tests. The pretest asked the students to write a paragraph comparing and contrasting their life now and five years ago, and the post-test topic was an argumentative paragraph on the writer's favorite type of art and why it is the best in their opinion. These topics were chosen based on the English essay writing course content and also the usual in-class writing tasks. As the participants were taking the Writing 1 course at the B.A. level, the main course objective was to instruct them to compose well-structured, coherent paragraphs; hence, assessing them by means of paragraph writing was in line with the goals set for this writing course. Also, the effect of addressing different writing genres was minimized by employing an analytic rubric, which mainly focuses on grammar and linguistic aspects rather than genre-based differences. As Imbler et al. (2023) argue, language learners are generally familiar with typical writing genres (e.g., comparison/contrast, argumentative, etc.) since they have been exposed to them frequently, and assessing their writings based on an analytic rather than holistic rubric makes the results rather objective to the genre addressed.

In order to grade the participants' paragraphs in both pre- and post-test, Jacobs et al.'s (1981) writing scale was employed (see the Appendix). This rating rubric is an analytical scoring technique that considers a set of criteria for scoring an essay (Weigle, 2002). These criteria consist of five specific dimensions: content, organization, vocabulary use, language use, and mechanics. The multi-dimensional nature of this scoring rubric makes it a successful scoring system, as it evaluates various aspects of a written text (Brown & Bailey, 1984). Throughout the course, learners were made aware of these criteria, as the writing tasks they were required to complete during the semester were structured based on these benchmarks.

Concerning the reliability of the tests, as Lou (2015) points out, a writing assessment task is believed to be reliable if it measures the same task by different raters consistently. Therefore, to assess the reliability of the writing tasks in the pre- and post-test, the first researcher (i.e., the teacher) first piloted the tests by giving them to a similar group of 120 students, and checked the consistency of the scores allocated by two different raters by calculating the interrater reliability. These raters were two experienced English teachers who have been teaching various English courses, including writing, for over 7 years and have developed a reputation for being knowledgeable, engaging, and supportive teachers in the English department of Farhangian University. The degree of agreement between the two raters was calculated to be 0.748, which is acceptable (McHugh, 2012). Then, the actual pre- and post-tests were scored using the same rubrics by the raters trained by the teacher. Regarding the face and content validity of the pre- and post-test, the raters confirmed that they appeared to be appropriate, and that the content of the writing prompts provided, as well as the required writing tasks, were in line with the content covered by the Writing 1 course syllabus and materials.

### *Data Collection Procedure*

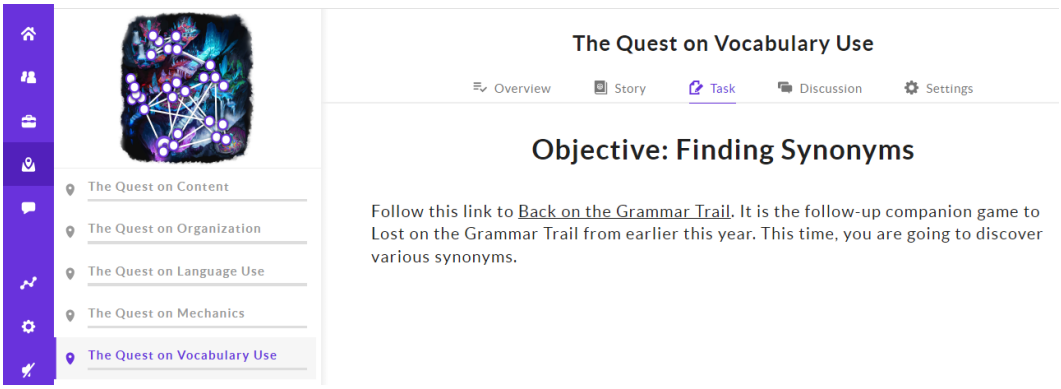
First, the participants took the DIALANG test to make sure that the sample is statistically homogeneous. Then the students took the pretest, which was to write a paragraph comparing and contrasting their life now and five years ago; and their writings were graded based on Jacobs et al.'s (1981) rubric, which included the following criteria: content, organization, vocabulary use, language use, and mechanics.

To apply gamification to the writing course, the Classcraft platform was employed, which offers a wide range of gamification elements and features. Classcraft is a digital role-playing gamified environment, which was launched in 2014, and was designed specifically to gamify classroom management (Sanchez et al., 2017). Its goal is to assimilate the classroom into a role-playing game, acknowledging the players' positive behaviors, such as attending class, helping classmates, being on time, etc., by granting special powers to them. As the players gain various powers, they can use them both individually and collaboratively in their teams. Therefore, the teamwork and collaboration, as well as the competition among the learners, and their urge to unlock new powers, make them more motivated and alleviate their stress to become more engaged in L2 writing practice.

The virtual classes were created on the Classcraft platform and held for a duration of 16 weeks (a university semester), for two hours a week. The next step was to organize the quests in the course for the first and second experimental groups. Classcraft allows free accounts to set up to six objectives for each quest. The quest objectives in this study required the student to submit the files of their assignments to complete the tasks. Figure 1 indicates a task within a devised quest created in Classcraft.

**Figure 1.**

*Task related to the objective "Finding Synonyms" within The Quest on Vocabulary Use*




As demonstrated in Figure 1, this particular task with the objective of "Finding Synonyms" required the students to complete a series of game-like exercises (provided in the embedded link) in order to develop their vocabulary repertoire. Figure 2, below, shows an instance of the vocabulary exercises the learners would confront by following the link inserted in the task instructions.

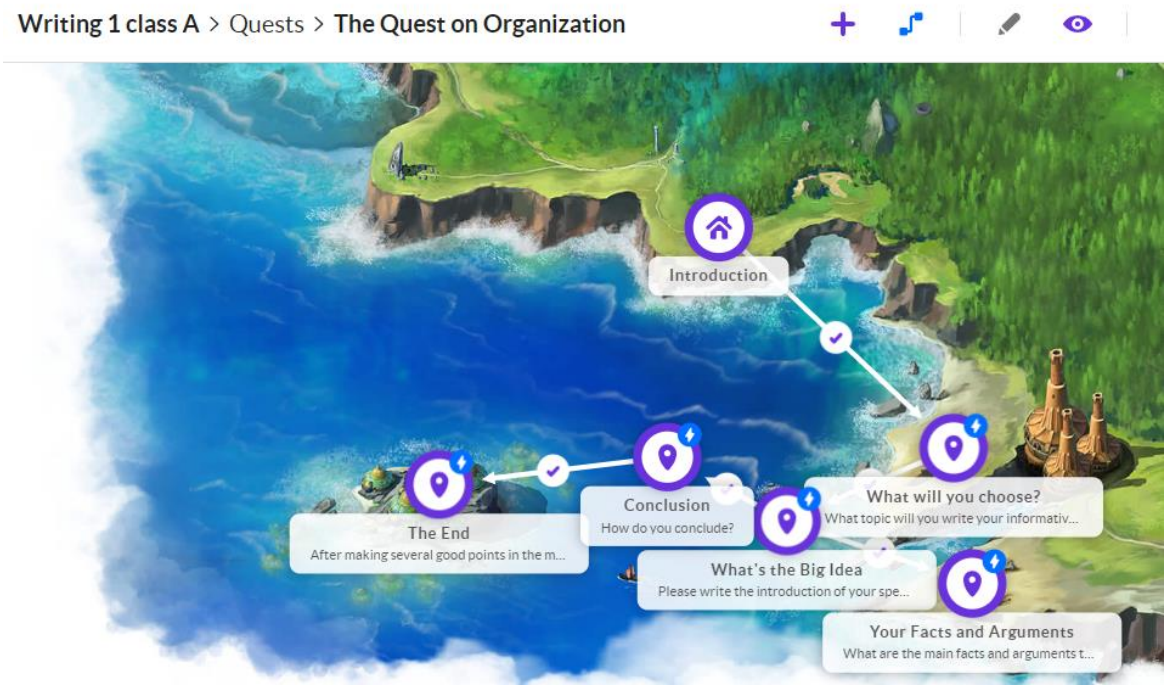
**Figure 2.**

*Example of a vocabulary exercise included in a task related to the objective "Finding Synonyms" within the Quest on Vocabulary Use*


Overall, five quests were devised, each concerning one of the criteria in the rubric, namely content, organization, vocabulary use, language use and mechanics; and each quest contained a set of objectives, which were the subcomponents of each criterion in the rubric. For instance, for the criterion of mechanics, the students were required to complete a quest which was composed of different objectives focused on spelling, capitalization, punctuation, and paragraphing. Figure 3 depicts a quest on the organization criterion containing four constituent objectives, as well as the beginning and ending points. Through this quest, the students were required to compose a coherent well-structured paragraph in several steps. Once all the objectives were set, a *Quest End* was added to conclude the quest and reward students for completing the entire quest.

**Figure 3.**

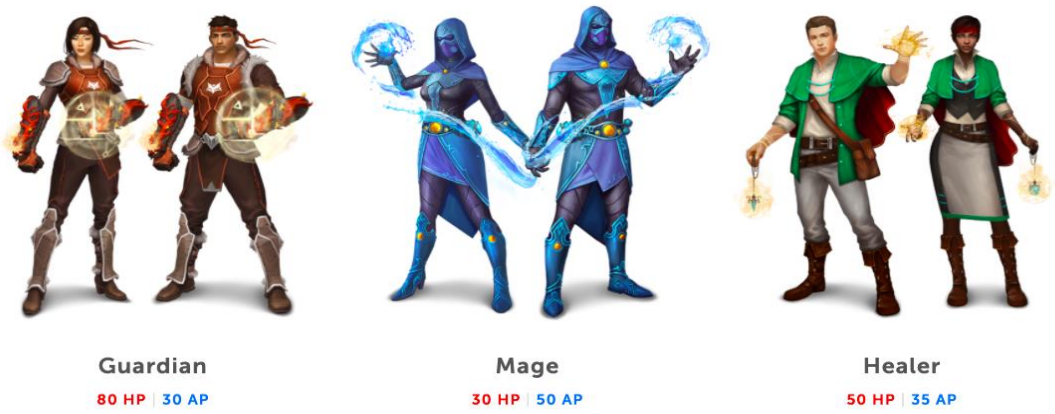
*Quest on organization criterion containing 4 objectives requiring students to compose a coherent organized paragraph*




After devising the quests, the students in the first and second experimental groups were trained to use Classcraft as a gamified educational platform. The teacher dedicated the whole first session to the introduction of this platform to the students in EG1 and EG2, walking them through its different parts and facilities. As they learned how to create their own accounts in Classcraft, the teacher added them either manually, via Google Classroom, or by giving students a code to join the class. The participants could create their own avatars, and they were each given the chance to decide if they wanted their avatar to be a *Warrior*, *Mage*, or *Healer* (see Figure 4), and to build their own identity in the gamified environment accordingly.

**Figure 4.**

*Examples of avatars in Classcraft*



Once all the students in the first experimental group were registered and had created their own avatars, they were randomly assigned to teams in order to start collaboration on completing the tasks in each quest. As there were three distinct types of characters available to choose from for the students, the optimum combination for a team was to include at least one role from each type (i.e., one *Mage*, one *Warrior* and one *Healer*). Thus, the students in the first experimental group were divided into groups of three. Since this group was supposed to do the tasks collaboratively, the writing assignments were completed via Google Docs, enabling them to work on their writing projects together. (The link was embedded in their Classcraft environment.) They could submit their assignments via the discussion part (similar to a forum), devised in their virtual classes in Classcraft, where all their classmates could see their answers and comment on them in the first experimental group, where they benefitted from their teammates' collaboration, as well. Also, the teacher regularly checked and gave feedback on the students' assignments in all groups. As the second experimental group were required to compose their writings individually, they were just added to the virtual class created in Classcraft without forming any teams. The participants in this group were required to complete the tasks individually; therefore, they could submit their assignments directly via the Classcraft platform, where they received feedback from the teacher. They did the tasks in each quest individually, as opposed to the first experimental group, who progressed through the quests as a team. The students in the third treatment group, who did the tasks collaboratively without the gamification element, used Google Docs to do their writing assignments. Finally, the control group, who experienced a writing course without either collaboration or gamification elements, submitted their writing assignments in Word or PDF format via the Farhangian University LMS. The students in different groups were not aware of the participants in the other groups, as there were no interactions between them.

Overall, the students in all groups produced five paragraphs, each written on a new topic selected from a variety of subjects drawn from their course materials; they then received feedback on their writings from the teacher. The teacher allocated grades to each assignment. In the case of the first and second experimental groups, the grades could then be converted into points by using a grade converter feature, or as it is called in this platform, *Treasures of Tavuros*, which transforms the results on quizzes, assignments, or exams into *Experience Points* rewards for the students. In Classcraft, *Experience Points* (XPs) are a measure of a student's progress and achievements. They are awarded for completing tasks, participating in class activities, and demonstrating positive behavior. Accumulating experience points allows students to level up, unlocking new powers and abilities within the platform, motivating and engaging them in their learning journey.

After completing the tasks, the participants in all groups took the post-test, which required them to write a paragraph on their favorite type of art and why it was the best in their opinion. All the learners took the post-test individually, as the purpose of the study was to see how their individual writing performance was affected by the treatment. Finally, the scores of the participants in all treatment and control groups were analyzed and compared to see whether there had been a significant change in their individual English writing performance.

### *Data Analysis*

The data gathered through the first phase of data collection (i.e., the DIALANG test), were analyzed regarding normality of distribution. Likewise, the scores that learners gained in the pre- and post-test were examined to see if they met the normality of data conditions. Afterwards, following other preliminary checks such as linearity, homogeneity of variances, and homogeneity regression slopes, a two-way ANCOVA was conducted, with pretest scores as

covariate, to see if the treatment and control groups had performed significantly differently from each other.

## Results

Initially, a group of 180 students were tested and the results of the DIALANG English placement test were used to choose a homogenous sample of 120 students consisting of both male and female participants in terms of their English proficiency out of the larger sample. Their scores were automatically calculated out of 1000 and the students whose scores were below 200 and over 600 were excluded from the study. Further, to figure out whether the dataset related to the DIALANG scores was normally distributed, Kolmogorov-Smirnov and Shapiro-Wilk tests were conducted, with the results suggesting that the DIALANG scores were normally distributed, and that the participants comprised a homogenous group of students.

After groupings the participants, all took the pretest, and at the end of the semester with their assigned treatments, all participated in the post-test. The descriptive statistics of the participants in the four groups (including three experimental and one control groups) are demonstrated in Table 2, indicating the numbers, gender, and proficiency level of the participants, as well as the mean and standard deviation of the pre- and post-test scores (rated out of 100) in each group.

**Table 2**.

*Descriptive statistics of pretest and post-test scores of the 3 experimental groups and the control group*

| | | | | | Pretest Scores | | Post-test Scores | |
|---|---|---|---|---|---|---|---|---|
| Groups | N | Male + Female | Age range | English Proficiency level | Mean | Std. Deviation | Mean | Std. Deviation |
| EG1 | 30 | 14 + 16 | 19 - 23 | Intermediate | 40.98 | 13.83 | 55.32 | 13.17 |
| EG2 | 30 | 13 + 17 | 19 - 23 | Intermediate | 44.16 | 12.88 | 51.74 | 13.05 |
| EG3 | 30 | 12 + 18 | 19 - 23 | Intermediate | 43.99 | 11.42 | 47.21 | 11.35 |
| CG | 30 | 12 + 18 | 19 - 23 | Intermediate | 40.56 | 12.74 | 43.50 | 12.95 |

EG = Experimental Group
CG = Control Group

Once the pre-test and post-test writings were rated, the data were checked via normality tests, with the results indicating normality of distribution. Furthermore, the values of skewness and kurtosis related to the scores fell within an acceptable range according to George and Mallery (2024) (i.e., between -2 and +2)., thus confirming the normality of the data, without including any outliers and extremes. Therefore, it was concluded that the scores of the participants in all groups in both the pretest and the post-test were normally distributed.

To answer the three research questions, a two-way ANCOVA was conducted. It is worth mentioning that the assumptions required to run this analysis on the obtained dataset were first examined. The assumptions of normality, linearity and reliable measurement of the covariate (i.e., the pretest scores) and the dependent variable (i.e., the post-test scores) were all checked and ensured for all four groups. The results of Levene's test of homogeneity of variances indicated that there were no significant differences between the variances of the different groups and therefore the assumption of the homogeneity of the variances across groups was met. There was also no interaction between the independent variables (i.e., gamification and collaboration factors) and the covariate, and the assumption of homogeneity of regression slopes was not

violated. Thus, the dataset was analyzed by running a two-way ANCOVA, as indicated in Table 3, with the post-test scores as the dependent variable, gamification and collaboration as the independent variables, and the pretest scores as the covariate.

**Table 3**.

*Results of two-way ANCOVA for the post-test scores of the participants compared in terms of gamification and collaboration factors in different groups*

| Source | Type III Sum of Squares | df | Mean Square | F | Sig. | Partial Eta Squared | Observed Power [b] |
|---|---|---|---|---|---|---|---|
| Corrected Model | 2715.195 [a] | 4 | 678.799 | 4.276 | .003 | .129 | .919 |
| Intercept | 16390.790 | 1 | 16390.790 | 103.248 | .000 | .473 | 1.000 |
| Pretest Scores | 311.669 | 1 | 311.669 | 1.963 | .164 | .017 | .285 |
| Gamification | 1999.602 | 1 | 1999.602 | 12.596 | .001 | .099 | .941 |
| Collaboration | 401.676 | 1 | 401.676 | 2.530 | .114 | .022 | .351 |
| Gamification * Collaboration | 5.066 | 1 | 5.066 | .032 | .859 | .000 | .054 |
| Error | 18256.514 | 115 | 158.752 | | | | |
| Total | 314301.207 | 120 | | | | | |
| Corrected Total | 20971.709 | 119 | | | | | |

a. R Squared = .129 (Adjusted R Squared = .099)
b. Computed using alpha = .05

Regarding the first research question, i.e., the effect of using a gamified platform on the individual L2 writing performance of Iranian TEFL students, Table 3 indicates that the participants using the Classcraft (i.e., the first and second experimental groups) collectively outperformed the participants who did not experience gamification, as there was a statistically significant difference: $F(1, 115) = 12.596$; $p = .001 < .05$). However, concerning the second research question, i.e., the effect of collaborative writing on individual L2 writing performance, there was no statistical difference between the groups: $F(1, 115) = 2.530$; $p = .114$; therefore, collaborative writing had no impact on the results. Similarly, regarding the third research question, i.e., whether there was an interaction between gamification and collaborative writing, as Table 3 illustrates, the interaction effect was not statistically significant: $F(1, 115) = .032$, $p = .859$.

In recent years, merely reporting the level of significance is considered insufficient; therefore, a look at the effect sizes and confidence intervals also seems in order: As seen in Table 3, the effect size (i.e., the partial eta squared, $\eta^2$) for the presence of gamification was .099, indicating a medium-to-large effect (Cohen, 1988). Also, based on the confidence interval reported in Table 4, the difference for the population means (i.e., with/without gamification) could vary between 3.6 and 12.7 points out of 100 (i.e., the maximum score) 95% of the time, which seems to indicate a moderate effect size. Furthermore, the width of the 95% confidence interval [3.608, 12.721] is less than twice the mean difference (i.e., $9.113 < 2 * 8.164$), which makes it quite narrow (Field, 2018, p. 241). This indicates that the difference for the population means for those who did or did not experience gamification via Classcraft is a real difference, and further research is not very likely to yield different results (Larson-Hall & Plonsky, 2015).

**Table 4**.

*Pairwise comparisons of the independent factors: Gamification and collaborati*

| (I) Groups | (J) Groups | Mean Difference (I-J) | Std. Error | Sig.[b] | 95% Confidence Interval for Difference[b] Lower Bound | Upper Bound |
|---|---|---|---|---|---|---|
| Gamification is present | Gamification is absent | 8.164* | 2.300 | .001 | 3.608 | 12.721 |
| Collaboration is present | Collaboration is absent | 3.659 | 2.300 | .114 | -.897 | 8.216 |

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

Contrary to gamification, neither collaboration nor the interaction effect (i.e., gamification*collaboration) yielded statistically significant results (Table 3), and the effect sizes of both (.022 and .000, respectively) were small (Cohen, 1988). However, as seen in Table 4, the width of the 95% confidence interval for the difference for the population means working collaboratively or individually was nearly 2.5 times the mean difference (i.e., 9.113 > 2 * 3.659), making it a relatively wide interval (Field, 2018) and the results more open to question. Thus, while the lower and higher bounds of the 95% confidence interval for differences do include zero, the wide interval and the relative smallness of the lower bound indicate that more research is possibly warranted with regard to the impact of collaborative writing.

## Discussion

The research questions of this study aimed to investigate whether gamifying an L2 writing course, combined with collaborative writing, can have a significant impact on the L2 writing performance of Iranian university students in TEFL, and whether there is a significant interaction effect between gamification and collaboration.

Regarding the first research question, the results revealed that students who utilized the gamified platform (regardless of whether they used it individually or collaboratively) outperformed other learners in terms of their improvement on L2 writing post-test scores, with a medium-to-large effect size. These findings are in line with previous research proving gamification to be an effective factor in developing the students' writing skills (Ocriciano, 2016; Sánchez, 2024). A number of studies in this regard have suggested that gamification can help alleviate the anxiety and the affective barriers usually associated with writing practices, especially in an academic context (e.g., Laffey, 2022; Yavuz et al., 2020). In addition, similar to the results of this study, previous studies have observed that the academic writing performance of students increases through gamification (e.g., Tantawi et al., 2018; Zhihao & Zhonggen, 2022).

As for the impact of collaborative writing on the individual L2 writing performance of the students (i.e., the second research question), the results of the two-way ANCOVA analysis

showed that the collaboration factor did not significantly affect the scores of the participants, regardless of whether or not they were using gamification. There are, in fact, contradictory results suggested by different studies conducted on the effectiveness of collaborative writing. While an increasing body of research insists on the advantages of implementing collaborative writing, as it fosters interaction between the learners and in effect increases the learners' motivation to write (e.g., Slavin, 2011), there are other studies which have reported the ineffectiveness of collaborative writing in their findings (e.g., Dobao, 2012; Strauss & U, 2007; Zabihi et al., 2013). Previous research findings have already demonstrated that learners in competitive conditions (e.g., gamified contexts) were more likely to adopt performance-oriented goals, which is more often accompanied by better short-term performance, while the collaborative mode requires longer periods of practice to exert significant effects (Li et al., 2024).

From another angle, as Kinnear (2021) contends, learning performance in educational settings is usually measured by exams, and the results more often reflect the individual performance rather than group performance. Thus, collaboration in groups might not necessarily affect the learning performance measured individually. In the present study, both the pre- and the post-test measured the individual L2 writing performance of the participants, and did not take collaborative performance into consideration, as it was not the variable examined in this particular context. In addition, Wiethof et al. (2021) found that "while group competition motivated some participants, it did not affect others at all" (p. 54). They argued that as collaboration among peers is of a more qualitative nature compared with gamification, "one needs to give participants enough time to come up with qualitatively valuable contributions" (p. 54) and that is quite different from the immediacy of the impact that gamification exerts on the learners' individual performance. Furthermore, Olsen et al. (2014) claimed that collaborative learning had positive effects on gaining conceptual knowledge but not procedural knowledge. When collaborating on conceptual problem-solving steps, the learners are usually urged to talk to each other and provide mutual explanations, which can lead to improved learning outcomes in a conceptual post-test in comparison to individual learning. However, when students are engaged in procedural problem-solving, they often do not provide mutual elaborations. Rather, they take turns in solving the different problem-solving steps, individually. In other words, the differences in the learning materials may have possibly triggered different types of collaborative behavior that are not equally effective for promoting student learning. In this study, the learners were required to submit the writing tasks collaboratively in the first and third experimental groups, but the type of collaboration among them was more of a procedural nature, as the quest tasks were devised in a step-by-step format. On the other hand, some researchers have discussed the question of equal contribution in the practice of collaborative writing. Strauss and U (2007) observe that it is not uncommon for the learners with better writing performances to take on the responsibility of producing the final writing piece. Therefore, making each learner write individually in a gamified environment (and possibly making the individuals more competitive in the process) might help them improve more in comparison with the collaborative mode (Ho et al, 2022).

Concerning the interaction effect between gamification and collaboration (i.e., the third research question), the findings of the two-way ANCOVA indicated that there was no significant interaction observed between these two factors. This finding is consistent with the limited body

of research examining both gamification and collaboration in different contexts (e.g., Guo et al, 2024; Ly, 2021). For instance, Ho et al. (2022) have stated that in their study "peer collaboration did not moderate the effectiveness of gamification in learning as no subgroup differences were found between collaborative games and non-collaborative games" (p. 3833). They observed that the impact of games that were both collaborative and competitive was not different from that of games that were merely competitive. They attributed this to the fact that gamification directly affects the individual performance while collaboration among peers requires a dynamic coordination among the individuals that might not necessarily be manifested in the learner's individual performance.

## Conclusion

Gamification can offer valuable opportunities to educators as well as students in L2 learning. If implemented appropriately, a gamified learning environment can create a stress-free atmosphere in which learners can experience a mode of learning and progress with higher levels of motivation, interest, and engagement (Reynolds & Taylor, 2020). The utilization of Classcraft for EFL students in the present study further provides confirmation of the effectiveness of gamification for learning writing, and incorporating gamified elements into writing exercises appears to be a promising approach to increase student engagement and improve learning outcomes. On the other hand, it would seem that investing in collaborative writing is not particularly helpful for improving individual writing. There may still be merit, however, in providing collaborative writing tasks in EFL situations, such as making the task more engaging and motivating for the learners, or providing them with practice opportunities in preparation for real-life collaborative writing. It is suggested that further research might be useful in this regard.

## References

Albarracin, D., Johnson, B. & Zanna, M. (2005). Attitudes: Introduction and scope. In D. Albarracin, B. Johnson, & M. Zanna (Eds), *The handbook of attitudes* (vol. 1, pp. 1–18). New York, NY: Routledge.

Anunpattana, P., Khalid, M., Iida, H., & Inchamnan, W. (2021). Capturing potential impact of challenge-based gamification on gamified quizzing in the classroom. *Heliyon, 7*(12).

Aubrey, S. (2022). Dynamic engagement in second language computer-mediated collaborative writing tasks: Does communication mode matter?. *Studies in Second Language Learning and Teaching*, *12*(1), 59–86.

Aydin, Z., & Yildiz, S. (2014). Using wikis to promote collaborative EFL writing. *Language Learning & Technology*, *18*(1), 160–180.

Brown, J. D., & Bailey, K. M. (1984). A categorical instrument for scoring second language writing skills. *Language Learning, 34*(4), 21–38.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cronk, M. (2012). Using gamification to increase student engagement and participation in class discussion. In T. Amiel & B. Wilson (Eds.), *Proceedings of EdMedia 2012--World conference on educational media and technology* (pp. 311-315). Denver, Colorado, USA: Association for the Advancement of Computing in Education (AACE).

Dao, P. (2020). Effect of interaction strategy instruction on learner engagement in peer interaction. *System*, *91*, 1–13.

Deterding, S. (2012). Gamification: Designing for motivation. *Interactions*, *19*(4), 14-17.

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). NY, USA: ACM.

DIALANG. (2024). Retrieved from https://dialangweb.lancaster.ac.uk/.

Dobao, A. F. (2012). Collaborative writing tasks in the L2 classroom: Comparing group, pair and individual work. *Journal of Second Language Writing, 21*, 40-58.

European Commission. (2018). Common European Framework of Reference for Languages: Learning, teaching, assessment. Retrieved from https://www.coe.int/en/web/common-european-framework-reference-languages/home.

Field, A. (2018). Discovering statistics using IBM SPSS statistics (5th ed.). Sage Publications Ltd.

George, D., & Mallery, P. (2024). IBM SPSS Statistics 29 Step by Step: A Simple Guide and Reference (18th ed.). New York: Routledge.

Guo, K., Zhong, Y., Zainuddin, Z., & Chu, S. (2024). Applying game-related methods in the writing classroom: A scoping review. *Education and Information Technologies,* **29**, 4481–4504.

Hashemi, S. A., Naderi, E., Shariatmadari, A., Naraghi, M. S., & Mehrabi, M. (2010). Science production in the Iranian Educational system by the use of critical thinking. *International Journal of Instruction, 3*(1), 61–76.

Ho, J. CS., Hung, Y. S. & Kwan, L. YY. (2022). The impact of peer competition and collaboration on gamified learning performance in educational settings: A Meta-analytical study. *Education and Information Technologies,* 27, 3833–3866.

Imbler, A. C., Clark, S. K., Young, T. A., & Feinauer, E. (2023). Teaching second-grade students to write science expository text: Does a holistic or analytic rubric provide more meaningful results? *Assessing Writing*, *55*(Complete). https://doi.org/10.1016/j.asw.2022.100676.

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfiel, V. F., & Hughey, J. B. (1981). *Testing ESL composition: A practical approach*. Newbury House.

Jeon, J. (2022). Exploring a self-directed interactive app for informal EFL learning: A self-determination theory perspective. *Education and Information Technologies*, 1–21.

Kalogiannakis, M., Papadakis, S., & Zourmpakis, A.-I. (2021). Gamification in science education: A systematic review of the literature. *Education Sciences*, *11*(1), 22. MDPI AG. Retrieved from http://dx.doi.org/10.3390/educsci11010022.

Kessler, G. (2009). Student-initiated attention to form in wiki-based collaborative writing. *Language Learning & Technology*, *13*(1), 79–95.

Kinnear, G. (2021). Two-stage collaborative exams have little impact on subsequent exam performance in undergraduate mathematics. *International Journal of Research in Undergraduate Mathematics Education, 7*(1), 33-60.

Laffey, D. (2022). Gamification and EFL writing: Effects on student motivation. *English Language & Literature Teaching, 28*(1), 25-44.

Larson-Hall, J. & Plonsky, L. (2015), Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, *65*(S1), 127-159.

Lee, L. (2004). Learners' perspectives on networked collaborative interaction with native speakers of Spanish in the US. *Language Learning & Technology*, *8*(1), 83–100.

Li, L., Hew, K.F. & Du, J. (2024). Gamification enhances student intrinsic motivation, perceptions of autonomy and relatedness, but minimal impact on competency: A meta-analysis and systematic review. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-023-10337-7.

Liu, G., 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*, 1-18.

Lopres, J., Lactam, J., Sia, J., Placencia Jr, M., Alili Jr, A., Labaria, J., Boloron, R., Ching, R., Alsaraireh, M., Capulso, L., & Reyes-Chua, E. (2023). A model of collaborative writing technique to enhance students' writing fluency. *World Journal of English Language, 13*(3), 227-238.

Lundstrom, K., & Baker, W. (2009). To give is better than to receive: The benefits of reviewing to the reviewer. *Journal of Second Language Writing*, *18*(1), 30-43.

Ly, Q. C. (2021). Using gamification to teach and engage students in the act of summary writing. *Journal of Media Literacy Education, 13*(3), 111-122.

Marandi, S. S., & Seyyedrezaie, M. S. (2017). The multi-course comparison of the effectiveness of two EFL writing environments: Google drive versus face-to-face on Iranian EFL learners' writing performance and writing apprehension. *CALL-EJ, 18*(1), 9–21.

Mazlan, C. A. N., Abdullah, M. H., Sulong, M. A., Abas, A., Ramdan, M. R., Safian, A. R., Syariff, M., & Fuad, D. R. (2023). Exploring the integration of bite-sized learning: A

scoping review of research in education and related disciplines. *Contemporary Educational Technology, 15*(4), ep468. DOI: https://doi.org/10.30935/cedtech/13622.

McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276-282.

Memari, M. (2021). Effects of critical thinking on writing cause and effect essays by Iranian EFL learners. *Research in English Language Pedagogy, 9*(2), 489 – 506.

Morris, B. J., Croker, S., Zimmerman, C., Gill, D., & Romig, C. (2013). Gaming science: The "Gamification" of scientific thinking. *Frontiers in psychology, 4*, Article 607. DOI: 10.3389/fpsyg.2013.00607.

Ocriciano, M. (2016). IELTS writing: A gamification journey. *Cambridge English: Research Notes, 64,* 31-38.

Olsen, J. K., Belenky, D. M., Aleven, V., & Rummel, N. (2014b). Collaboration on procedural problems may support conceptual knowledge more than you may think. In *Proceedings 16 of the 3rd Workshop on Intelligent Support for Learning in Groups at the 12th International Conference on Intelligent Tutoring Systems*.

Reynolds, E., & Taylor, B. (2020). Kahoot!: EFL instructors' implementation experiences and impacts on students' vocabulary knowledge. *Computer-Assisted Language Learning Electronic Journal, 21*(2), 70-92.

Rezai, A. (2022). Cultivating Iranian IELTS candidates' writing skills through online peer feedback: A mixed-methods inquiry. *Education Research International*, vol. 2022. Retrieved from https://doi.org/10.1155/2022/6577979.

Rezaei, M., & Jafari, M. (2014). Investigating the levels, types, and causes of writing anxiety among Iranian EFL students: A mixed-method design. *Procedia-Social and Behavioral Sciences*, *98,* 1545 – 1554.

Rincon-Flores, E. G., Mena, J., & López-Camacho, E. (2022). Gamification as a teaching method to improve performance and motivation in tertiary education during COVID-19: A research study from Mexico. *Education Sciences, 12*(49). Retrieved from https://doi.org/10.3390/educsci12010049.

Salimi, A., & Zanganeh, H. (2022). The effect of gamification on vocabulary learning (learning English as a second language) among the fifth-grade elementary school students. *Technology of Education Journal (TEJ)*, *16*(4), 723-734. DOI: 10.22061/tej.2022.8550.2686.

Sánchez, X. C., Caicedo Alvarez, K.I., & Parra-Gavilánez, L. F. (2024). The effectiveness of gamification to enhance writing skills in english as a foreign language contexts. In: M. E., Auer, U. R., Cukierman, E. Vendrell Vidal, & E. Tovar Caro (eds.) *Towards a hybrid, flexible and socially engaged higher education, ICL 2023: Lecture notes in networks and systems* (vol. 911, pp. 366–371). Springer, Cham.

Sanchez, E., Young, S., & Jouneau-Sion, C. (2017). Classcraft: From gamification to ludicization of classroom management. *Education and Information Technologies*, *22*(2), 497 – 513.

Sari, M., & Erten, İ. H. (2021). Developing a digital placement test for English as a foreign language: A validity study. *Computer Assisted Language Learning*, 1-27.

Schmid, R., Pauli, C. & Petko, D. (2023). Examining the use of digital technology in schools with a school-wide approach to personalized learning. *Educational Technology Research and Development, 71*, 367–390.

Slavin, R. E. (2011). Instruction based on cooperative learning. In R. E. Mayer & P. A. Alexander (Eds.), *Handbook of research on learning and instruction*, (pp. 344-360). Taylor & Francis.

Storch, N. (2011). Collaborative writing in L2 contexts: Processes, outcomes, and future directions. *Annual Review of Applied Linguistics*, 31, 275–288.

Storch, N. (2013). *Collaborative writing in L2 classrooms.* Bristol, UK: Multilingual Matters.

Stott, A. & Neustaedter, C. (2013). *Analysis of gamification in education, technical report 2013-0422-01.* Connections Lab, Simon Fraser University, Surrey, BC: Canada.

Strauss, P., & U, A. (2007). Group assessments: Dilemmas facing lecturers in multicultural tertiary classrooms. *Higher Education Research & Development, 26*(2), 147-161.

Talaue, F. G. (2023). Dissonance in generative AI use among student writers: How should curriculum managers respond?. *E3S Web of Conferences. 426*. https://doi.org/10.1051/e3sconf/202342601058

Tanis, C. (2020). The seven principles of online learning: Feedback from faculty and alumni on its importance for teaching and learning. *Research in Learning Technology*, *28*. https://doi.org/10.25304/rlt.v28.2319.

Tantawi, M. E., Sadaf, S., & AlHumaid, J. (2018). Using gamification to develop academic writing skills in dental undergraduate students. European Journal of Dental Education, 22, 15-22.

Vameghshahi, B. M., & Ghonsooly, B. (2023). A video game-based paragraph writing instruction vs. teacher-based writing instruction: Examining L2 learners' perceptions through dynamic assessment. *Teaching English as a Second Language Quarterly (Formerly Journal of Teaching Language Skills), 42*(1), 121-146.

Verswijvelen, M., Sosa, R., & Martini, N. (2020) Designing game-inspired narratives for learning, In Boess, S., Cheung, M. & Cain, R. (Eds.), *Synergy - DRS International Conference 2020*, (pp. 11-14). [Held online in August 2020]. https://doi.org/10.21606/drs.2020.294.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, *31*(2), 57–71.

Weigle, S. C. (2002). *Assessing writing*. Cambridge University Press.

Wiethof, C., Tavanapour, N., & Bittner, E. (2021). Designing and evaluating a collaborative writing process with gamification elements: Toward a framework for gamifying collaboration processes. *AIS Transactions on Human-Computer Interaction, 13*(1), 38–61.

Wigglesworth, G., & Storch, N. (2009). Pair versus individual writing: Effects on fluency, complexity and accuracy. *Language Testing, 26*(3), 445-466.

Yavuz, F., Özdemir, E., & Celik, O. (2020). The effect of online gamification on EFL learners' writing anxiety levels: A process-based approach. *World Journal on Educational Technology, 12*, 62 –70.

Yu, S., & Lee, I. (2016). Peer feedback in second language writing (2005–2014). *Language Teaching*, *49*(4), 461–493.

Zabihi, R., Rezazadeh, M. & Vahid Dastjerdi, H. (2013). Creativity and narrative writing in L2 classrooms: Comparing individual and paired task performance. *Bellaterra Journal of Teaching & Learning Language & Literature, 6*(3), 29-46.

Zhihao, Z., & Zhonggen, Y. (2022). The impact of gamification on the time-limited writing performance of English majors. Education Research International, 2022, 1 – 11.

Zhou, L., Yu, J., & Shi, Y. (2017). Learning as adventure: An app designed with gamification elements to facilitate language learning; *Proceedings of the International Conference on HCI in Business, Government, and Organizations* (9–14 July 2017). Vancouver, BC, Canada.

## Biodata

Zahra Golesorkhi is currently a PhD candidate in TEFL at Alzahra University. She obtained her Bachelor's degree in English Language and Literature from Shahid Beheshti University, followed by a Master's degree in TEFL (Teaching English as a Foreign Language) from the same university. She has been teaching at Farhangian University, Shahid Bahonar Branch since 2019. Her main areas of interest include CALL (Computer-Assisted Language Learning), Gamification, and GBLL (Game-Based Language Learning).

Seyyedeh Susan Marandi is a professor in the Department of English at Alzahra University. She is known as "mother of CALL" in Iran, where she has, among other achievements, established the first CALL courses and master's program in CALL, as well as the first and only Iranian CALL Research Center. Prof. Marandi has published widely in national and international journals.

## Appendix: Rubrics Used to Assess Students' Writing

Table 1. *Analytical scoring rubric* (Jacobs et al., 1981 cited in Weigle, 2002, p. 115-116).

| ASPECT | SCORE | LEVEL/ CRITERIA |
|---|---|---|
| CONTENT | 30-27 | EXCELLENT TO VERY GOOD: knowledgeable • substantive • thorough development of thesis • relevant to assigned topic |
| | 26-22 | GOOD TO AVERAGE: some knowledge of subject • adequate range • limited development of thesis • mostly relevant to the topic, but lacks detail |
| | 21-17 | FAIR TO POOR: limited knowledge of subject • little substance • inadequate development of topic |
| | 16-13 | VERY POOR: does not show knowledge of subject • non-substantive • not pertinent • OR not enough to evaluate |
| ORGANIZATION | 20-18 | EXCELLENT TO VERY GOOD: fluent expression • ideas clearly stated/ supported • succinct • well-organized • logical sequencing • cohesive |
| | 17-14 | GOOD TO AVERAGE: somewhat choppy • loosely organized but main ideas stand out • limited support • logical but incomplete sequencing |
| | 13-10 | FAIR TO POOR: non-fluent • ideas confused or disconnected • lacks logical sequencing and development |
| | 9-7 | VERY POOR: does not communicate • no organization • OR not enough to evaluate |
| VOCABULARY | 20-18 | EXCELLENT TO VERY GOOD: sophisticated range • effective word/ idiom choice and usage • word form mastery • appropriate register |
| | 17-14 | GOOD TO AVERAGE: adequate range • occasional errors of word/ idiom form, choice, usage *but meaning not obscured* |
| | 13-10 | FAIR TO POOR: limited range • frequent errors of word/ idiom form, choice, usage •*meaning confused or obscured* |
| | 9-7 | VERY POOR: essential translation • little knowledge of English vocabulary, idioms, word form • OR not enough to evaluate |
| LANGUAGE USE | 25-22 | EXCELLENT TO VERY GOOD: effective complex constructions • few errors of agreement, tense, number, word order/ function, articles, pronouns, prepositions |
| | 21-18 | GOOD TO AVERAGE: effective but simple constructions • minor problems in complex constructions • several errors of agreement, tense, number, word order/ function, articles, pronouns, prepositions *but meaning seldom obscured* |
| | 17-11 | FAIR TO POOR: major problems in simple/ complex constructions • frequent errors of negation, agreement, tense, number, word order/ function, articles, pronouns, prepositions and/ or fragments, run-ons, deletions •*meaning confused or obscured* |
| | 10-5 | VERY POOR: virtually no mastery of sentence construction rules • dominated by errors • does not communicate • OR not enough to evaluate |
| MECHANICS | 5 | EXCELLENT TO VERY GOOD: demonstrates mastery of conventions • few errors of spelling, punctuation, capitalization, paragraphing |
| | 4 | GOOD TO AVERAGE: occasional errors of spelling, punctuation, capitalization, paragraphing *but meaning not obscured* |
| | 3 | FAIR TO POOR: frequent errors of spelling, punctuation, capitalization, paragraphing • poor handwriting •*meaning confused or obscured* |
| | 2 | VERY POOR: no mastery of conventions • dominated by errors of spelling, punctuation, capitalization, paragraphing • handwriting illegible • OR not enough to evaluate |