The Impact of Quiz Mode on Students' Learning Achievement: A Gamified e-Quiz Study

https://doi.org/10.3991/ijet.v18i12.34155

Kyung-Mi O(✉)
Dongduk Women's University, Seoul, Korea
kmo@dongduk.ac.kr

Abstract—This study investigated the impact of quiz mode on university students' academic achievement in an online ESOL theory course as well as their perceptions of the gamified e-quiz mode. Forty-two students at a women's university in Seoul participated in the study for 12 weeks. Utilizing a crossover design (AB/BA), this study compared the academic achievements of Group A (n=22) and Group B (n=20) after experiencing both gamified and conventional e-quiz interventions. Group A was treated with gamified e-quizzes for an initial three weeks, had a three-week break, and was then provided with conventional e-quizzes for three weeks. Group B received the same treatments in reverse order: taking the conventional e-quizzes first, having a washout break, and then taking gamified e-quizzes. Three tests—one pre-test and two post-tests—were developed and used to measure the participants' academic achievement. Two parallel survey forms were used after each intervention to measure the students' perceptions. The data analyses of the test results indicated that the students' academic achievements after the gamified e-quizzes were not significantly different from those after the conventional e-quizzes. In addition, the survey data illustrated, except for some data sets indicating more favorable perceptions toward gamified e-quizzes, that the participants of both groups viewed the two modes of quizzes mostly positively. Moreover, more participants in both groups chose gamified e-quizzes over conventional ones, describing the gamified e-quizzes as enjoyable, motivating, and low-anxious experiences.

Keywords—gamification, gamified e-quiz, learning achievement, class engagement, crossover design

1 Introduction

Educators have been in pursuit of an engaging online learning environment that keeps the attention of digital natives, and gamification has been recognized for its potential to hold the learners' engagement with enjoyable, motivating, and productive learning environments. Thus, gamification has been a budding topic in the field of education for the last decade and has bloomed along with the many online course offerings provided during the pandemic.

Previous gamification studies have discussed several issues, including motivation, attitudes, satisfaction, and learning achievement [1, 2, 3, 4, 5]. Gamification has been widely adopted in diverse areas such as mathematics, medical science, business, engineering, and language education [6, 7, 8, 9, 10].

Despite the recent popularity, gamification studies on learning achievement in courses on theories of second language learning are limited in number, and corresponding empirical data are in need. Given that, this study was designed to address these needs in two aspects.

First, this study examined the impact of quiz mode on learning achievement by comparing the interventions of gamified and conventional e-quizzes. As an empirical study conducted on students majoring in English at a Korean university, this study aimed to compare the impact of gamified and conventional e-quizzes on the students' academic achievement in a course on theories of second language learning.

Second, the study employed a crossover model so that all the participants could experience both experimental and control groups in balance. Specifically, this study employed the AB/BA model, allowing the two groups to be both the experimental and control groups, to experience both gamified e-quizzes and conventional e-quizzes for their learning achievement and class engagement. Since each group worked as its control case, the model could reduce inter-subject variability from the comparison between groups, and the effect of covariates could also be decreased. By employing the crossover design, this study was expected to have more power and statistical efficacy and obtain more reliable data from the participants exposed to both quizzes in balance.

Accordingly, this paper aims to answer the following research questions:

1. Do the students' learning achievements differ between the gamified and conventional e-quiz groups?
2. Is there a difference in the students' perceptions toward gamified e-quizzes and conventional e-quizzes for their class engagement?
3. After experiencing both gamified and conventional e-quizzes, which one do the students prefer?

2 Literature review

Gamification has been defined by many research studies [11, 12, 13]. A widely accepted definition is "the use of game design elements in non-game contexts" ([13], p. 10). As the word "game" suggests, gamification has often been perceived to "create enjoyable, fun, and motivating learning experiences for learners" ([14], p. 936) and has been applied in both digital and non-digital environments [15]. In particular, digital gamification has been a rising research topic in the studies of technology-enhanced learning with its potential to achieve positive learning outcomes through enjoyable online learning processes [15].

Thus, gamification has been implemented in various fields, such as nursing, medical science, information technology, and education, to serve educational needs. The application of this nascent approach in each area has been investigated in several aspects, including students' learning motivation [16], self-regulated learning [17], cooperative

learning [18], and class engagement [19]. Much gamification research has been released on student motivation [1, 20, 21] and class engagement [3, 19, 22]. Yet, relatively little research has examined the relationship between learning outcomes and gamification [23].

The relative scarcity of gamification studies displaying its impact on learning outcomes seems partly related to the difficulty in directly measuring learning achievement. As one of the few studies examining the relationship between gamification and learning achievement, Li et al. [24] explored the impact of a gamified e-quiz system called *Reading Battle* on students' reading experiences. *Reading Battle* was a gamified e-quiz bank created to improve primary students' reading skills. The data analyses indicated the gamified e-quiz experience helped the students to develop reading habits and motivation. Most students reflected that gamification was helpful for them to form their reading habits by increasing reading frequency and motivation and widening the book selection range. The study was a good case in point, illustrating how gamification could facilitate primary students' reading motivation and autonomous reading practices. Although the researchers did not directly measure the students' reading performances, they noted that gamification has the potential to increase learning achievement after examining the students' self-reports about the gamified e-quiz after the intervention in their study.

Unlike the previous study [24], Palaniappan et al. [25] chose to measure the learning outcomes for their gamification study directly. By employing a single group of 29 university students, the researchers measured the students' subject knowledge of programming language by administering pre-and post-tests to investigate the impact of a 4-week gamification intervention. The researchers found that the students' academic performances improved significantly after the gamification intervention. The researchers also used a questionnaire to assess the students' self-directed learning before and after the gamification, and they found a meaningful positive change after the intervention. The researchers illustrated that the learners could control and pursue their learning through gamification activities.

The study [25] is a good example, detailing how gamified e-quizzes and assessments can be implemented in a university course and what students can expect after the activities. The study's sample size was relatively small, with 29 students, and the research design was simple, using a single group pre-and post-test design. Since the single-group performance measurement was conducted before and after the intervention, the group served as its own control group, allowing the study to be replicable and economical.

With a more rigorous research design, Ahmed et al. [26] examined the effects of gamification on Iranian EFL learners' idiomatic knowledge. Using an experimental design, the researchers found that gamification helped the learners improve their idiomatic knowledge. The researchers noted that the strengths of gamification, including enjoyable learning, prompt feedback, and lowered anxiety, were the reasons for the students' positive perceptions [26].

The study [26] is judged to have good internal validity since it employed randomization for the two-group assignment. Due to the random assignment, the researchers could observe how the two groups' idiomatic knowledge changed from the pre-test to the post-test, attributing the difference observed in the post-test not to the possible difference between the two groups before the experiment but to the gamification intervention.

In another recent study [23] into experimental gamification, the researchers were interested in the impact of gamification on the timed English writing performances of 76 EFL university students. For the study, two groups were asked to complete a 35-minute writing composition task every week for five weeks. To examine whether the gamified group showed noticeable improvement in their timed English writing, the researchers employed a mixed-method approach combining quantitative and qualitative data. The quantitative analyses indicated that the gamified group in week five outperformed their performance in week one in all four aspects of fluency, accuracy, syntax, and morphology. Unlike the noticeable improvement observed in the gamified group, the control group displayed a slight improvement in week five from week one.

Their digital data gathered from the platform (i.e., learning times, task point completion, homework completion, and participation in the discussion) also indicated that the gamified group's participation exceeded that of the control group, suggesting that gamification played a part in the learners' class engagement. The researchers' analyses also revealed that the students' writing performance peaked in week three when the gamification reward reached the highest level. Through detailed descriptions, the researchers claimed that the gamification impact on students' learning motivation decreased after reaching its peak. Based on the findings, the researchers confirmed [27], suggesting that the beneficial impact of gamification is likely to be short-term, often labeled as a "novelty effect." As seen in the study, as the novelty effect decreased over time, the learning motivation was also expected to be reduced.

Thus, the study [23] seems to be meaningful in that it examined the impact of gamification on students' timed writing performances, which had not been discussed previously. However, as the researchers' mentioned their study's weakness, as gamification elements included points, rankings, badges, and grades, it is difficult to identify which factors influenced the writing performances. Moreover, since the basic premise of gamification is to provide students with an enjoyable learning environment, removing the grade from the elements seems more appropriate.

In summary, a few studies have provided valuable insight into the use of gamification. Unlike earlier studies on the subject, often with small scales and single groups, recent studies have used a more rigorous design, employing a control group, random assignment, and a mixed-method approach. Still, more systematic studies on gamification are in need. In particular, there is a lack of empirical research comparing the efficacy of gamified and conventional e-quizzes on academic achievement. This comparison is in demand because the typical pre-test/post-test control group design with gamified quiz treatment often makes it difficult to determine if the result is due to the gamification intervention or simply due to the quiz treatment. Thus, as a small step toward rigorous gamification research, this study did not rely on the typical experimental design (e.g., [23, 28]) but utilized a crossover design so the participants could work as their own control. Therefore, this study was designed to examine students' academic achievements and perceptions after using gamified e-quizzes and conventional online quizzes as a crossover gamification experiment.

3 Methodology

3.1 Research design

This study employed a pre-test/post-test experimental crossover design. A crossover study, also known as a crossover trial, usually uses the AB/BA model [29]. In the model, one treatment group of a crossover study is provided with treatment A followed by treatment B. In contrast, the other group is provided with treatment B followed by treatment A for the so-called "balance" between the two groups [29]. The crossover study is efficient because the participants' responses to treatment A are contrasted with the reactions to treatment B, removing the participant variation [29, 30]. Thus, using the AB/BA model, two types of e-quizzes were applied to assess the educational effects on the university students taking an online ESOL theory course in this study.

3.2 Participants and setting

The participants were employed from one of the researcher's courses in the English Department in the fall semester of 2021 at a women's university in Seoul, Korea. As one of the ESOL theory courses offered by the department, the course is designed to provide first-year students majoring in English with information about research findings on teaching English and theoretical views on second language acquisition. The course is designed with two 75-minute sessions per week over a 15-week semester, and this study was started from week 4, allowing this project a 12-week study.

For the two 75-minute sessions, the researcher/instructor uploaded two 60-minute video lessons for the students to view weekly, followed by two 10-minute quizzes and two 5-minute Q&A activities. Of the two 75-minute sessions, however, 15 minutes of one session was spent synchronously with half of the participants for live gamification, and 15 minutes of the other session was used on the rest half students for synchronous conventional e-quiz activities. Out of 50 registered students, 44 volunteered for the study and were randomly assigned to two groups. However, two students dropped out after the random assignment, so 42 students participated in this study and were divided into Group A (n=22) and Group B (n=20). Students in both groups experienced both gamified and conventional e-quiz interventions and responded to Surveys I and II. All the participants were females in their twenties, and their majors were English. All human subject research approvals were obtained from the university's institutional review board (Protocol Number DDWU2103-01).

3.3 Instruments

Quizzes and feedback. For the gamified and conventional e-quiz interventions, twelve e-quizzes were constructed to assist students in checking their general understanding and comprehension of critical concepts covered in the class video lessons based on the textbook, *How Languages are Learned* [31]. Specifically, each quiz was designed to assess a student's comprehension of each session. Therefore, each student took two quizzes per week, meaning six quizzes for each 3-week intervention and a total of twelve for the current research design with two 3-week interventions.

Each quiz was composed of 10 multiple-choice and true/false items to be solved within 10 minutes, allowing one minute per item.

Conventional e-quizzes were implemented using the school's LMS platform and used the same quiz questions as the gamified e-quizzes. Gamified e-quizzes were created using the Quizizz platform (https://quizizz.com/), which is widely employed for motivating students through gamification [32]. The platform is known for various gamified elements, including avatars, leaderboards, memes, music, themes, and time constraints for gamifying learning activities [22]. The gamified e-quizzes allowed the students to assess their understanding of the content within a game-like environment, enjoy live gamification by competing with their classmates, and receive prompt feedback depending on their answers (see Figure 1).

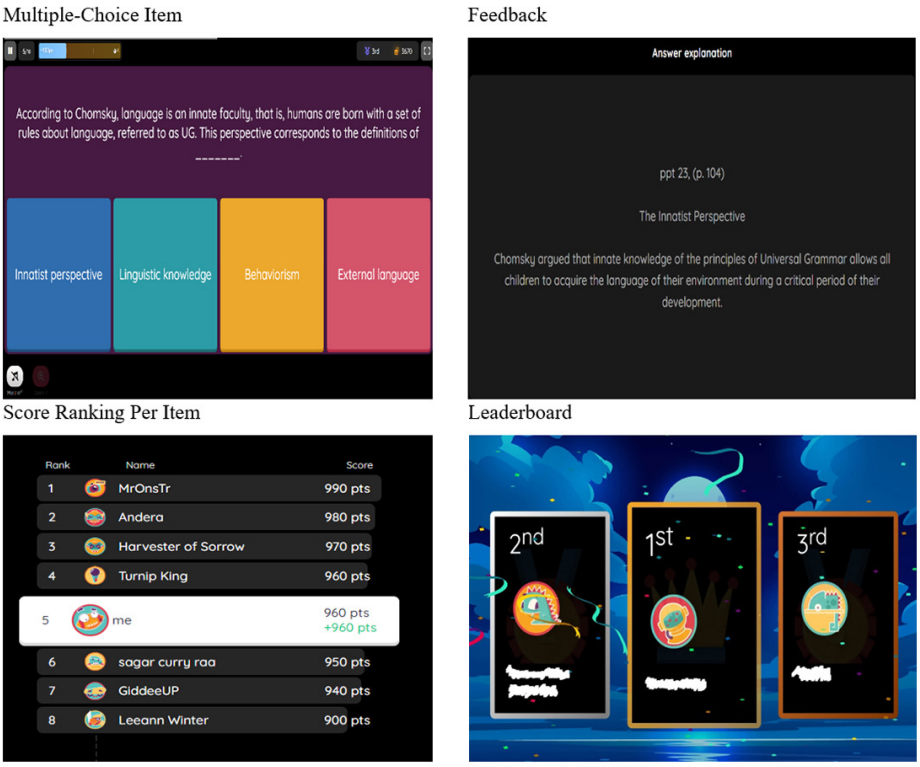

Fig. 1. Gamified e-quiz samples

Each question in both the gamified and conventional platforms had the same written feedback with an answer or a content summary. In addition to the written feedback, video clips of the researcher explaining each quiz item were also provided for the two groups. The feedback video clips were primarily the researcher/instructor's explanations of the correct answers to the items. Lastly, the Naver Café (https://section.cafe.naver.com/ca-fe/) messaging system was available around the clock for effective Q&A with the students. This system provided instant alerts to the instructor.

Test instruments. Three online test instruments (Pre-test I and Post-tests I and II) were developed for the study. Pre-test I was used twice for measuring the difference between the two groups before the first intervention and retesting the difference before the second intervention. Thus, it needed to be short and compact to retain the students' interest and participation in the study rather than risk having them quit in the middle. Pre-test I was developed with 20 true/false items, each worth two points, based on the content covered in the students' coursebook. The 20 items required students to apply and synthesize ESOL theories without relying on simple factual recalls [33, 34] from the course content or direct connections to Post-tests I and II. The test items were revised through the analyses of two English testing experts for content validity. The modified items were piloted with five students comparable to the participants, and the calculated value (r=0.81) using the KR-21 formula was judged to be reliable enough to be used.

Post-tests I and II (midterm and final examinations) were used to measure academic achievement after the first and second interventions. Each post-test consisted of 10 multiple-choice and 10 true-false items, each worth two points for a total score of 40. As the analyses focused on comparing the two groups' performances on the two post-tests, using identical parameters for both pre-and post-tests was unnecessary. As a mid-term and a final exam, each post-test included items to measure various concepts and theories covered in the course textbook as assigned in the syllabus. Since the test items on the midterm and final exams were to be reflected in the students' grades, pilot tests were impossible. In place of pilot tests, the test item statistics from a pool of previous achievement test questions were used to adjust item difficulty. For the content validity of the tests, the two English testing experts reviewed them, and the items were revised based on the review. Both Post-test I (KR-20 coefficient r=0.54) and Post-test II (KR-20 coefficient r=0.66) were marginally reliable. Cronbach's alpha value of 0.7 or above has been widely reported to be reliable. However, the alpha value of around 0.5 has also been used as a satisfactory reliability value in much research [35]. The reason is that an instrument such as an achievement or diagnostic test measuring conceptual knowledge with loosely coherent constructs cannot render high reliability [36]. A low Cronbach's alpha or KR-20 would be reasonable in many studies [37, 38, 39]. And a high-reliability value cannot guarantee the reliability of the test, but often it merely means there are redundant test items to be deleted [38]. Nonetheless, the reliability values of the instruments are marginal or at least modest. Thus, the interpretation drawn from the test data must be cautiously made.

Survey. Two parallel survey forms were designed with 18 Likert scale items and 7 open-ended questions. The 18 Likert scale items were adapted from [22], initially developed by [40]. The initial survey form consisted of 22 items on a seven-point Likert scale based on four cognitive, behavioral, emotional, and agentic engagement domains [40]. The scale was partly revised and used in [22] for their gamification study. The researchers [22] changed the original scale into 24 items on a five-point Likert scale measuring the exact four domains as [40]. Of the 24 items, except for six in the agentic domain, 18 were used with minimal revision for this study. The agentic domain was excluded since it was judged to be deviant from the rest of the items [41].

For the open-ended questions, two interview questions (#21 and 22) were used with minimal revision from [22], which were also adapted initially from [40]. Two questions were added, asking about the strengths and weaknesses of each quiz intervention.

Three additional questions were included in Survey II that asked the participants to compare the two quiz interventions after experiencing each.

For the scale's internal consistency, Cronbach's alpha value was tallied. The survey was reliable enough to be used since the calculated value was 0.92.

3.4 Procedure

This research project was conducted on an online ESOL theory course for 12 weeks (see Figure 2).

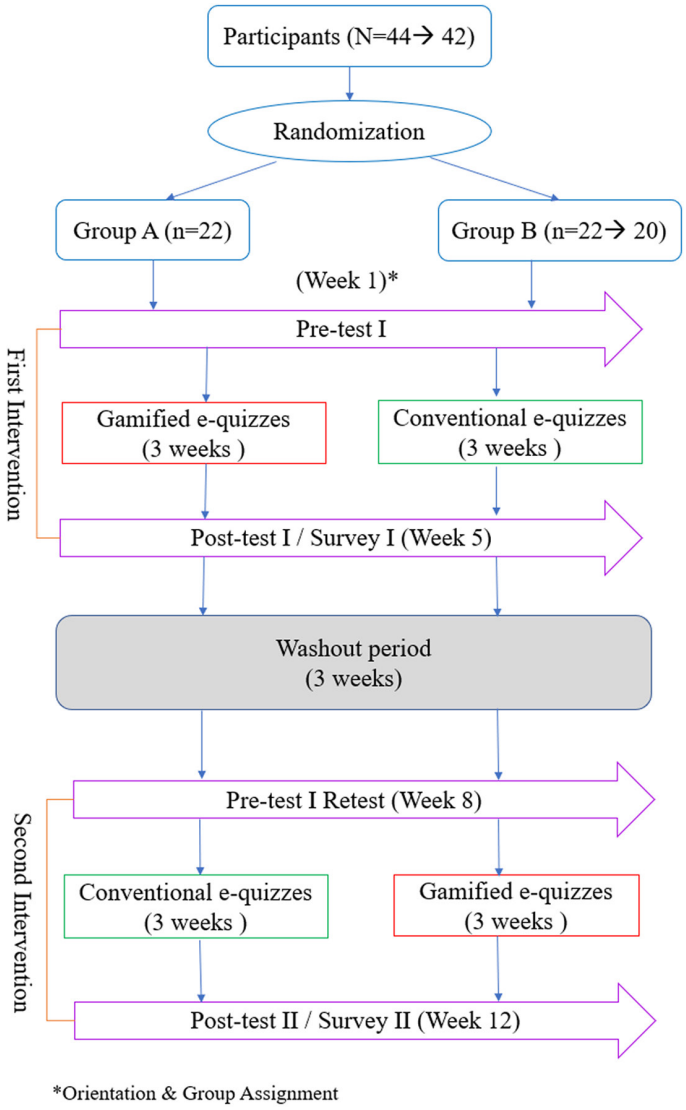

Fig. 2. Intervention procedure

At the initial stage, the participants were randomly assigned to either Group A or B. During the first session of week one, the students in both groups were introduced to the current project and evaluation criteria. Notably, the students were reminded that their scores from gamified and conventional e-quizzes would not be reflected in their grades but would be used only for their own learning. They were allowed to use any materials, including lecture notes and other quiz questions, to solve the two quiz modes. In the second session of week one, Pre-test I was administered online to the two groups for fifteen minutes. After the pre-test, the two groups of students were provided identical class content in the form of two 60-minute video lessons, each followed by an e-quiz designed to be a 10-minute self-assessment learning activity. Over three weeks, students in Group A completed six gamified e-quizzes, while students in Group B completed six conventional e-quizzes. At the end of the first intervention for each group, the respective surveys for the two modes and Post-test I (midterm examination) were provided, followed by a three-week break for washing out the previous intervention.

After the washout period, following [42] suggestion for additional baseline measurement to reduce the effect of random variation, Pre-test I was administered again in week eight at the beginning of the second intervention. Then, the two quiz modes were switched between the two groups. For the next three weeks, Group A was offered six conventional e-quizzes, and Group B was given six gamified e-quizzes. At the end of the second intervention, the respective surveys for the two modes and Post-test II (final examination) were implemented.

3.5 Data analyses

For analyzing the quantitative data, SPSS version 26 was used. Since this study had a small sample size, analyzing the data distribution was essential for choosing an appropriate statistical method. To that end, a Shapiro-Wilk test was performed, and the test result showed that some of the two groups' test results were not normally distributed. The data from Group A's Pre-test I ($W(22)=0.84$, $p=.002$) and Post-test I ($W(22)=0.86$, $p=.005$) departed significantly from normality.

In addition, the carryover effect from the crossover design was expected to be minimal for a few reasons. First, this study is not a medical study involving an acutely curable disease that can be cured after the first intervention. Additionally, the three-week washout period was included after the first intervention. Moreover, Pre-test I was reused to ensure no carryover effect following the washout period. Under the circumstances, for analyzing the non-parametric quantitative data of this study, analyses comparing the difference within and between the groups were conducted using the Wilcoxon signed-rank test and Mann-Whitney *U* test.

The Mann-Whitney *U* test was utilized for the Likert scale data to compare the data between the groups. The paired Wilcoxon signed-rank test was conducted to analyze within-group data. For all the statistical tests, including the Wilcoxon signed-rank test and the Mann-Whitney *U* test, $p < 0.05$ was deemed statistically significant. The qualitative data from the open-ended questions were analyzed using Microsoft Excel. Some responses in Korean were translated into English and quoted, and several answers in English were directly cited.

4 Results

4.1 The impact of test mode on learning achievement

The first research question asked if there was a difference in the students' learning achievements between the gamified and conventional e-quiz modes. The quantitative analyses showed that the learning achievements of the two groups did not display a significant difference between the two modes.

The descriptive statistics of the two groups' test performances from both pre-tests and post-tests are shown in Table 1. Both groups displayed means and medians of around 20 out of 40 on Pre-test I, Pre-test I Retest, and Post-test II.

Table 1. Descriptive data of the two groups over two interventions

	First Intervention		**Second Intervention**	
	Pre-Test I	**Post-Test I**	**Pre-Test I Retest**	**Post-Test II**
	Median (IQR)/ Mean (SD)	**Median (IQR)/ Mean (SD)**	**Median (IQR)/ Mean (SD)**	**Median (IQR)/ Mean (SD)**
Group A (n=22)	*22.00 (3.00)*	32.00 (5.00)	*21.00 (6.00)*	24.00 (11.00)
Sequence GC	*20.45 (5.24)*	31.36 (4.64)	*20.82 (5.75)*	22.82 (6.97)
Group B (n=20)	*21.00 (7.00)*	34.00 (6.00)	*21.00 (8.00)*	24.00 (13.00)
Sequence CG	*22.20 (4.20)*	32.70 (4.17)	*21.50 (4.25)*	24.60 (6.19)

First, to observe the impact of the e-quiz mode, the two groups' homogeneous achievements on the two pre-tests and the possibility of crossover effect were examined. The Mann-Whitney *U* test between the groups' scores on Pre-test I indicated the two groups did not display significant differences, *U* (N Group A=22, N Group B=20)=202.00, $z=-0.460$, $p=0.645$ ($p>0.05$). In addition, the two groups displayed insignificant results on their Pre-test I Retest, *U* (N Group A=22, N Group B=20)=212.50, $z=-0.191$, $p=0.849 > 0.05$. Thus, the Mann-Whitney *U* test results suggest that the two groups are homogeneous regarding the two pre-tests. The Wilcoxon signed-rank test result on the two pre-tests indicated no crossover effect from the first intervention. Specifically, there was no significant difference between the Pre-test I (*Mdn*=22.00) and the Pre-test I Retest (*Mdn*=21.00) of Group A, $z=-0.29$, $p=0.977 > 0.05$, $r=-0.004$; the Pre-test I (*Mdn*=21.00) and the Pre-test I Retest (*Mdn*=21.00) of Group B, $z=-0.714$, $p=0.475 > 0.05$, $r=-0.113$, indicating no crossover effect.

Second, the Wilcoxon signed-rank test was carried out to examine each group's within-group variance on the two post-tests (see Table 2). The Wilcoxon signed-rank test result of Group A's scores between Post-test I (*Mdn*=32.00) and Post-test II (*Mdn*=24.00), $z=-3.669$, $p=0.000 < 0.05$, $r=-0.553$ indicated a statistically significant difference between the two post-tests. A similar result was observed in Group B. The Wilcoxon signed-rank test of Group B's scores on Post-test I (*Mdn*=34.00) and Post-test II (*Mdn*=24.00), $z=-3.582$, $p=0.000 < 0.05$, $r=-0.566$, also displayed a significant variance between the two post-tests. As shown in Table 2, each group displayed similar within-group variance on the two post-tests, with significantly higher medians on Post-test I than Post-test II.

Table 2. Within group variance between post-tests I and II

Group	Median (IQR)		*z*-Value	*p*-Value
	Post-Test I	Post-Test II		
A (n=22) (GC)	32.00 (5.00)	24.00 (11.00)	–3.669	0.000*
B (n=20) (CG)	34.00 (6.00)	24.00 (13.00)	–3.582	0.000*

Note: **p* < 0.05/G: Gamified e-quiz/C: Conventional e-quiz.

The Mann-Whitney *U* test was performed to examine any differences between the two groups on the two post-tests (see Table 3). According to the Mann-Whitney *U* test between the groups, the two groups did not display significant differences on both Post-tests I and II. With both tests, the smallest *p*-value was observed on Post-test I, *U* ($N_{Group\ A}$=22, $N_{Group\ B}$=20)=183.50, *z*=–0.932, *p*=0.351 (*p*>0.05). In addition, the two groups displayed similar results on Post-test II, *U* ($N_{Group\ A}$=22, $N_{Group\ B}$=20)=192.00, *z*=–0.715, *p*=0.475>0.05. Thus, the Mann-Whitney *U* test results of the two groups suggested that the two groups were homogeneous regarding the two post-tests.

Table 3. The Mann-Whitney *U* test between the two groups

Test	Mean Rank		*U*	*z*	*p*-Value
	Group A (n=22)	Group B (n=20)			
Post-test I	19.84 (G)	23.33 (C)	183.50	–0.932	0.351
Post-test II	20.23 (C)	22.90 (G)	192.00	–0.715	0.475

Note: **p* < 0.05/G: Gamified e-quiz/C: Conventional e-quiz.

In sum, despite the Wilcoxon signed-rank test of each group displaying the students' significantly higher scores on Post-test I than Post-test II, according to the Mann-Whitney *U* test, the two groups did not demonstrate differences on Post-test I. Thus, the results showed that both groups performed equally well on Post-test I, confirming [28], suggesting the test impact had a more significant influence than the quiz mode.

Unlike the two groups' high scores on Post-test I, their scores on Post-test II were not as high. This result can be seen as confirming [23] and [28], indicating the novelty effect, a short-term impact of a new treatment that wanes away over time. Just as the test impact peaked in the third week of the experiment [23], after three weeks of the first intervention in the current study, the scores on Post-test II of both groups decreased. This difference between the two post-test scores might be interpreted as the test impact in both groups straying over time, as observed in the previous studies.

Nevertheless, the interpretation has to be made with caution. The two post-tests were designed to be parallel; however, since one was a midterm examination and the other was a final examination covering different academic contents, they were not identical. Thus, the current study's data are insufficient to confirm the novelty effect. In understanding the contexts, it seems appropriate to summarize that the quantitative analyses illustrated that the academic achievements of the two groups with gamified and conventional e-quizzes were not statistically different.

4.2 Student perceptions of gamified e-quizzes and conventional e-quizzes

The second research question inquired if there was a difference in the students' perceptions toward gamified e-quizzes and conventional e-quizzes for their class engagement. Except for only a few items, the quantitative data analyses did not show statistically significant differences in between-group variance and within-group variance. The students in both groups had mixed perspectives, with positive and neutral attitudes toward each quiz (see Figure 3).

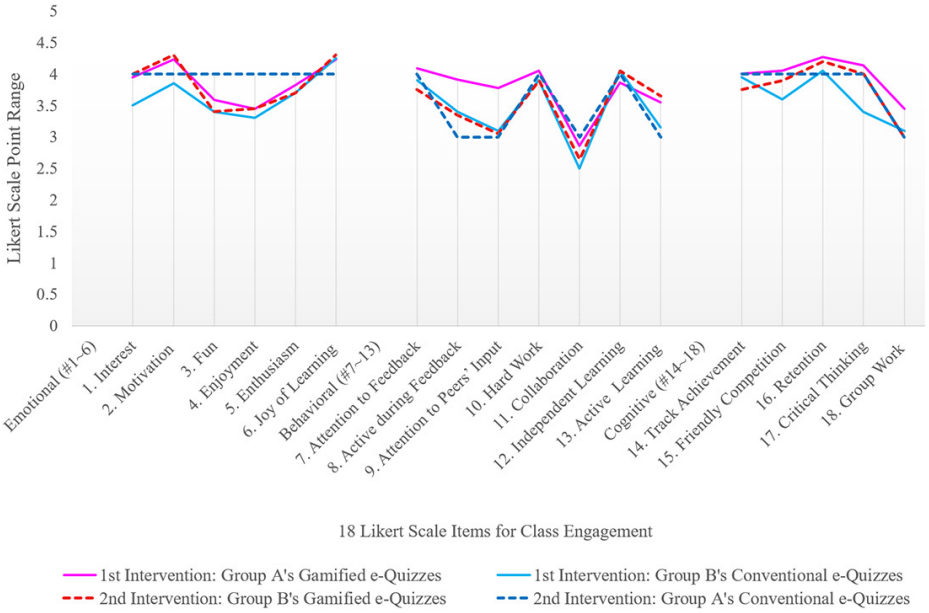

Fig. 3. Students' perceptions of each quiz mode for class engagement

The Mann-Whitney *U* test results rendered no statistically significant difference between the two groups except for the data related to the four items of the first intervention. To the four items (#1, 9, 15, & 17), Group A's responses to the gamified e-quizzes were more favorable than Group B's responses to the conventional e-quizzes with statistically significant differences ($p<.05$). However, the two groups did not render statistically meaningful results to the same items related to the second intervention ($p>0.05$). The data from the two groups were judged to be homogenous with the rest of the items. Students in both groups answered positively to 10 out of the 18 Likert scale items (#1, 2, 5, 6, 7, 10, 12, 14, 15, & 16) for both interventions. To the other eight, the two groups displayed neutral responses to item 11 for both interventions and responded with mixed reactions with neutral and positive answers to the other seven (#3, 4, 8, 9, 13, 17 & 18).

Regarding the within-group variance, according to the Wilcoxon signed-rank test results, the students in both groups did not display significant differences between the two interventions except for four cases. The data from the four (Group A's #9 &

Group B's #1, 2, & 17) revealed that the students in both groups provided more favorable responses to gamified e-quizzes with statistical significance.

The quantitative analyses of the Likert scale data suggested that the students in both groups perceived the two types of quizzes as mostly useful for their class engagement. The two data sets were similar except for a few cases with statistical significance from the Mann-Whitney *U* test and the Wilcoxon signed-rank test. The statistically significant results from both the Mann-Whitney *U* test (#1, 9, 15, & 17 in the first intervention) and the Wilcoxon signed-rank test (Group A's #9 & Group B's #1, 2, & 17) demonstrated that the students in both groups provided more favorable responses to gamified e-quizzes.

4.3 Students' preference after both interventions

After studying with both gamified and conventional e-quizzes, which quiz mode did the students prefer? The students responded to this question, as shown in Figure 4.

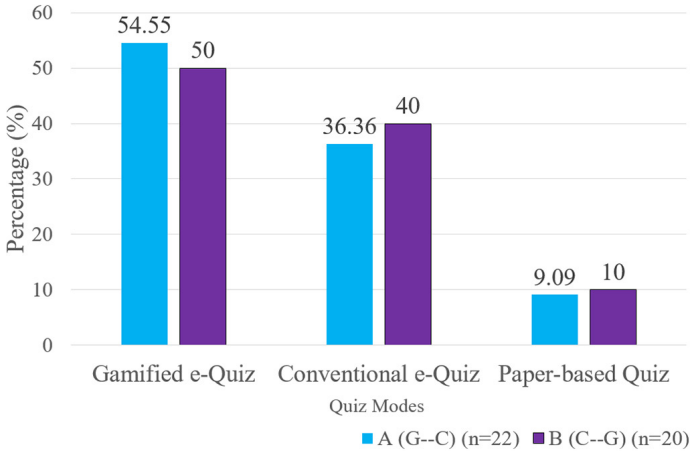

Fig. 4. Students' Preference after both gamified and conventional e-quizzes

The students in both groups indicated that they favored gamified e-quizzes over conventional e-quizzes. Paper-based quizzes were the least preferred. Twelve students (54.55%) from Group A (n=22) and ten students (50.00%) from Group B (n=20) chose gamified e-quizzes. Eight students (36.36%) from Group A and the same number of students (40.00%) from Group B chose conventional e-quizzes. Two students from each group indicated they still preferred paper-based quizzes to conventional online and gamified e-quizzes as they preferred writing or reading on paper.

As for their choosing either gamified or conventional e-quizzes, the students explained their selections with the following reasons (Figure 5).

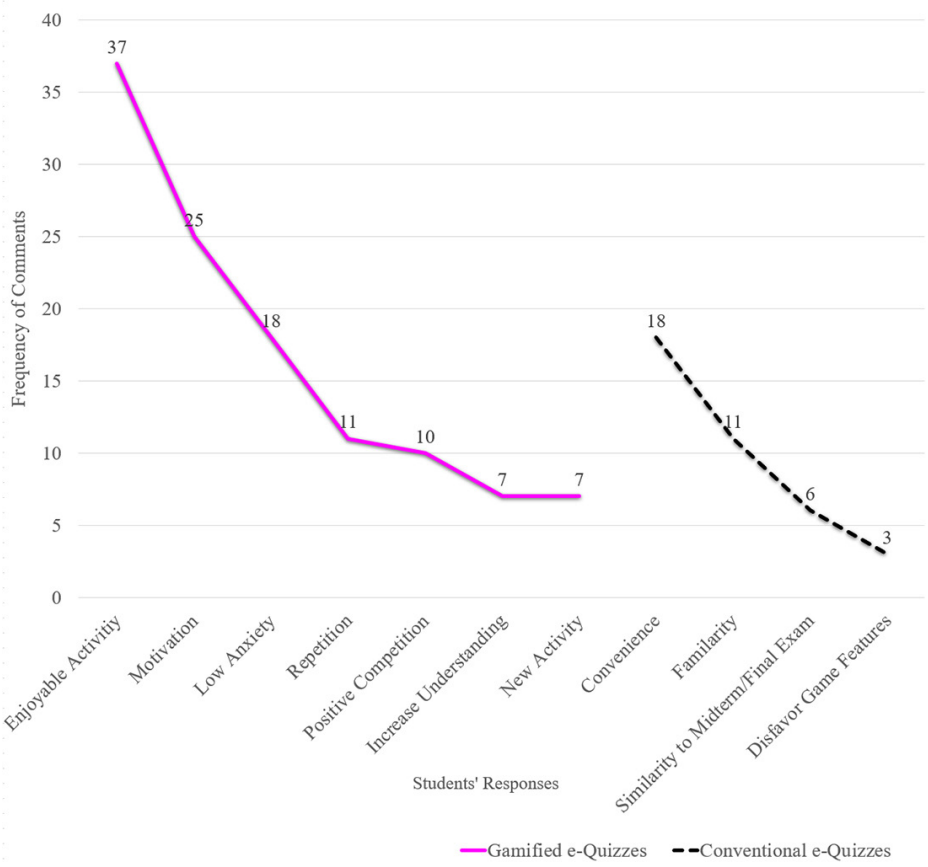

Fig. 5. Students' reasons for their preferences

In citing reasons for their preference for the gamified e-quizzes, most students chose the enjoyable activity, motivation, and low anxiety as their top three reasons. Among them, the enjoyable learning experience was the most prominent strength of gamified e-quizzes, and this was reinforced in the data collected from Item #25. To the question asking which form of the quizzes was more enjoyable, many students--Group A (86.36%) and Group B (80%)--chose the gamified e-quiz. The students from both groups elaborated on the enjoyable aspects of gamification:

A_ref.#21: Gamified e-quiz was more enjoyable because I felt the activity was a fun game rather than learning. It was just like a fun game.

A_ref.#4: It was more enjoyable because it increased the interest in learning, going beyond stereotypical learning and testing.

B_ref.#2: Gamified e-quiz because it makes studying more interesting.

B_ref.#14: Gamified e-quiz. It's more interesting.

The second most frequently mentioned reason was *motivation:*

A_ref.#7: A quiz in a game—adding points and solving problems—was more motivating.
B_ref.#19: It didn't make me feel pressured but encouraged me as a facilitator to keep going!

In addition to these two reasons, the gamified e-quiz mode was chosen because it allowed the students to feel *less anxious* compared to the conventional e-quiz mode:

B_ref.#9: The atmosphere of gamified e-quiz was more comfortable than the conventional one, so it was better for engagement.
A_ref.#10: It's studying but not in a typical study format. So I felt less stressed before starting.
B_ref.#10: I felt less pressure during the gamified e-quiz activities.
B_ref.#20: The former (gamified e-quiz) was more fun because I could play the game without feeling burdened.

Many students perceived the gamified e-quizzes as enjoyable, less anxious experiences, and motivating. However, some students in both groups still preferred the conventional e-quizzes for different reasons. In indicating a preference for the conventional e-quizzes, the dominant three reasons were convenience, familiarity, and test preparation. Specifically, the students chose the conventional e-quiz mode because they perceived accessing the quiz to be more convenient:

B_ref.#4: The conventional e-quiz was more convenient as I could access it through the university platform.
B_ref.#5: The conventional e-quiz. It was easily accessible.

Convenient access to the quizzes was the most frequently mentioned reason, and this was related to their familiarity with the platform:

B_ref.#16: I chose the conventional e-quiz. I'm not sure if this is because I prefer simple things, but the conventional e-quiz was easy to take and familiar to me.
A_ref.#22: I was more used to this form (the conventional e-quiz).

The familiarity with this quiz mode was partly related to the class platform provided by the university. Since most students used the university's platform daily during the pandemic, they were already familiar with conventional online quizzes. Moreover, most of them took summative tests in their classes utilizing the platform, and the students in this study addressed the point:

B_ref.#3: The conventional e-quiz uses a test format that we have continuously used for online classes (for years). So, I felt that I should take the quizzes. … because the quizzes were made in the test format, I thought I should take them. I wouldn't have taken the quizzes if I hadn't prepared for the formal (midterm and final) tests.
A_ref.#14: It (The conventional e-quiz) is similar to the actual tests (midterm and final exams).

Overall, despite some differences, the most preferred quiz mode of the two groups was the gamified e-quiz. Due to the current study's sample size and sampling method limitations, the data analyses cannot conclusively claim the students' dominant preference for gamified e-quizzes over the other two modes. However, this crossover study clearly illustrated that the most favored quiz mode for each group, regardless of the sequence of the intervention, was the gamified e-quiz. Thus, the current data analyses demonstrate that more students preferred the gamified e-quiz mode after experiencing both quiz modes.

5 Discussion and conclusion

This paper aimed to examine the efficiency of gamified e-quizzes on students' academic achievement by analyzing the difference between gamified and conventional e-quizzes. To that end, this study answered three research questions. To the first inquiry, asking if there was a difference in the students' learning achievements between the gamified and conventional e-quiz groups, the quantitative data analyses indicated that the learning achievements of the two groups were not significantly different.

The second question inquired if there was a difference in the students' perceptions toward the gamified and conventional e-quizzes for their class engagement. Aside from a few items revealing the two groups' more favorable attitudes toward the gamified e-quiz mode, the statistical analyses did not display significant differences. The participants in both groups had mixed perceptions with positive and neutral views toward both modes.

The last research question asked which mode of quiz the students preferred. After experiencing the two types of quizzes, more students in each group selected gamified e-quizzes over the conventional e-quizzes. As for their selection reasons, the students reported that they favored the gamified e-quiz mode because of the enjoyable and motivating experiences in low-anxious environments, confirming previous gamification research [26].

Although both groups perceived gamified e-quizzes more positively than conventional e-quizzes, Group B's preference toward the gamified e-quizzes was slightly lower than Group A's. This result appears to display the potential order effect because Group A was first exposed to the gamified e-quiz mode, and Group B was initially provided with the conventional e-quiz mode. Nonetheless, the issue of treatment order was dissolved since the study design used a counterbalancing technique within the frame of crossover design. By providing the gamified e-quiz mode in varying order, this study displayed that more students in each group, whether they experienced the gamified e-quizzes first or not, chose the gamified e-quiz mode over the conventional one.

The crossover design allowed the researcher to compare the same participants' performances after the two interventions, eliminating between-participant variability [29, 30]. Furthermore, through the crossover design, the participant could compare and select their preference after experiencing both modes of quizzes.

Despite the rigorous research design, the study has a few vulnerabilities. First, the study is limited in its sample selection and size. The class employed in the study was merely one of the researcher's classes. Only students who volunteered from a small number of registered students were used in the study, and two dropped out immediately after the random assignment. Thus, many aspects, such as the number of test items and the test designs, were considered to avoid losing participants.

Furthermore, since the class was an ESOL theory course with a strictly planned out syllabus, the test-retest technique for the two post-tests was unable to be used. Thus, each group's within-group analyses of academic achievements should be interpreted cautiously. More importantly, the reliabilities of the two post-tests were marginal. Several factors might have influenced the reliabilities. One of the factors is that each test item covered different concepts or theories, rendering little redundancy or repetition. And the contents measured in Post-test I might have been relatively easier than those in Post-test II. With all these issues, further studies are needed with more rigorous testing instruments administered to larger populations.

Nonetheless, through the systematic comparisons of between-group and within-group data, this study demonstrates that the gamified e-quiz mode can be used as the conventional e-quiz mode for students' academic achievement. The data also illustrate that both modes were perceived as mostly useful for class engagement. Thus, it seems possible to judge that the current data confirm the study [28], emphasizing the testing impact. As the study [28] suggested, the test usage seemed more meaningful for educational purposes rather than the specific quiz mode.

The current data illustrated that gamified e-quizzes could be as useful as conventional e-quizzes for students' learning achievement. Only the gamified e-quiz mode was favored over the conventional e-quiz in several survey items with statistically significant differences. Moreover, most students in each group chose the gamified e-quiz mode for an enjoyable learning experience. Furthermore, after experiencing both types of quizzes, more students from each group selected the gamified e-quiz mode over the conventional one. Therefore, this study is a case in point demonstrating that gamified e-quizzes are comparable to conventional ones with many benefits for students' academic achievement and class engagement.

6 References

[1] Ü. Çakiroğlu and S. Kiliç, "How to gamify?: Example scenarios for participation in synchronous online learning," E-Learning and Digital Media, vol. 15, no. 5, pp. 254–266, 2018. https://doi.org/10.1177/2042753018798166

[2] D. Codish and G. Ravid, "Academic course gamification: The art of perceived playfulness," Interdisciplinary Journal of e-Skills and Lifelong Learning, vol. 10, pp. 131–151, 2014. https://doi.org/10.28945/2066

[3] M. Leaning, "A study of the use of games and gamification to enhance student engagement, experience and achievement on a theory-based course of an undergraduate media degree," Journal of Media Practice, vol. 16, no. 2, pp. 155–179, 2015. https://doi.org/10.1080/14682753.2015.1041807

[4] J. Marti-Parreño, E. Méndez-Ibáñez, and A. Alonso-Arroyo, "The use of gamification in education: A bibliometric and text mining analysis." Journal of Computer Assisted Learning, vol. 32, no. 6, pp. 663–676, 2016, https://doi.org/10.1111/jcal.12161

[5] M. Sailer and M. Sailer, "Gamification of in-class activities in flipped classroom lectures," British Journal of Educational Technology, vol. 52, no. 1, pp. 75–90, 2021. https://doi.org/10.1111/bjet.12948

[6] A. A. A. Ahmed, M. Widodo, A. Komariah, I. Hassan, N. Sukmana, M. H. Ali, A. K. Abbas, and A. Rohi, "Assessing the effects of gamification on developing EFL learners' idiomatic knowledge: Do attitudinal factors contribute to the learning of the idioms with the games?," Educational Research International, vol. 2022, Article ID 2482570, 10 pages, 2022. https://doi.org/10.1155/2022/2482570

[7] H. Lajane, R. Gouifrane, R. Qaisar, F. Z. Noudmi, S. Lotfi, G. Chemsi, and M. Radid, "Formative e-assessment for Moroccan Polyvalent nurses training: Effects and challenges," International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 14, pp. 236–251, 2020. https://doi.org/10.3991/ijet.v15i14.13821

[8] V. Malvasi, J. Gil-Quintana, and E. Bocciolesi, "The projection of gamification and serious games in the learning of mathematics multi-case study of secondary schools in Italy," Mathematics, vol. 10, no. 3, pp. 336, 2022. https://doi.org/10.3390/math10030336

[9] M. Milanesi, S. Guercini, and A. Runfola, "Let's play! Gamification as a marketing tool to deliver a digital luxury experience." Electronic Commerce Research, vol. 2022. https://doi.org/10.1007/s10660-021-09529-1

[10] K. O., "A comparative study of gamified and conventional online quizzes," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 3, pp. 152–172, 2022. https://doi.org/10.3991/ijet.v17i03.26349

[11] K. M. Kapp, The gamification of learning and instruction: Game-based methods and strategies for training and education. San Francisco, CA: Pfeiffer, 2012. https://doi.org/10.1145/2207270.2211316

[12] K. M. Kapp, "2014 Reflections on gamification for learning," January 2015. [Online]. Available: https://karlkapp.com/2014-reflections-on-gamification-for-learning/ [Accessed July 22, 2022].

[13] S. Deterding, D. Dixon, R. Khaled, and L. Nacke, "From game design elements to gamefulness: Defining gamification," in Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, Tampere, Finland, 2011, pp. 9–15. https://doi.org/10.1145/2181037.2181040

[14] H. Dehghanzadeh, H. Fardanesh, J. Hatami, E. Talaee, and O. Noroozi, "Using gamification to support learning English as a second language: A systematic review," Computer Assisted Language Learning, vol. 34, no. 7, pp. 934–957, 2019. https://doi.org/10.1080/09588221.2019.1648298

[15] N. Sajinčič, A. Sandak, and A. Istenič, "Pre-service and in-service teachers' views on gamification," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 3, pp. 83–103, 2022. https://doi.org/10.3991/ijet.v17i03.26761

[16] M. D. Hanus and J. Fox, "Assessing the effects of gamification in the classroom: A longitudinal study on intrinsic motivation, social comparison, satisfaction, effort, and academic performance," Computers & Education, vol. 80, pp. 152–161, 2015. https://doi.org/10.1016/j.compedu.2014.08.019

[17] W. S. Alhalafawy and M. Z. Tawfiq Zaki, "How has gamification within digital platforms affected self-regulated learning skills during the COVID-19 pandemic? Mixed-methods research," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 6, pp. 123–151, 2022. https://doi.org/10.3991/ijet.v17i06.28885

[18] C. U. Bilgin and A. Gul, "Investigating the effectiveness of gamification on group cohesion, attitude, and academic achievement in collaborative learning environments," TechTrends, vol. 64, pp. 124–136, 2020. https://doi.org/10.1007/s11528-019-00442-x

[19] D. O. Göksün and G. Gürsoy, "Comparing success and engagement in gamified learning experiences via Kahoot and Quizizz," Computers & Education, vol. 135, pp. 15–29, 2019. https://doi.org/10.1016/j.compedu.2019.02.015

[20] N. V. Iaremenko, "Enhancing English language learners' motivation through online games," Information Technologies and Learning Tools, vol. 59, no. 3, pp. 126–133, 2017. https://doi.org/10.33407/itlt.v59i3.1606

[21] M. Sailer, J. U. Hense, S. K. Mayr, and H. Mandl, "How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction," Computers in Human Behavior, vol. 69, pp. 371–380, 2017. https://doi.org/10.1016/j.chb.2016.12.033

[22] Z. Zainuddin, M. Shujahat, H. Haruna, and S. Chu, "The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system," Computers & Education, vol. 145, Article ID 103729, 2020. https://doi.org/10.1016/j.compedu.2019.103729

[23] Z. Zhihao and Y. Zhonggen, "The impact of gamification on the time-limited writing performance of English majors," Educational Research International, vol. 2022, Article ID 4650166, 11 pages, 2022. https://doi.org/10.1155/2022/4650166

[24] X. Li, S. W. Mok, Y. Y. J. Cheng, and S. K. W. Chu, "An examination of a gamified e-quiz system in fostering students' reading habit, interest and ability," In Proceedings of the 81th Annual Meeting of Association for Information Science and Technology, Vancouver, Canada, 2018, vol. 55, pp. 290–299. https://doi.org/10.1002/pra2.2018.14505501032

[25] K. Palaniappan and N. M. Noor, "Gamification strategy to support self-directed learning in an online learning environment," International Journal of Emerging Technologies in Learning (iJET), vol. 17, no. 3, pp. 104–116, 2022. https://doi.org/10.3991/ijet.v17i03.27489

[26] A. A. A. Ahmed, M. Widodo, A. Komariah, I. Hassan, N. Sukmana, M. H. Ali, A. K. Abbas, and A. Rohi, "Assessing the effects of gamification on developing EFL Learners' idiomatic knowledge: Do attitudinal factors contribute to the learning of the idioms with the game?," Educational Research International, vol. 2022, Article ID 2482570, 10 pages, 2022. https://doi.org/10.1155/2022/2482570

[27] J. Koivisto and J. Hamari, "Demographic differences in perceived benefits from gamification," Computers in Human Behavior, vol. 35, pp. 179–188, 2014. https://doi.org/10.1016/j.chb.2014.03.007

[28] D. R. Sanchez, M. Langer, and R. Kaur, "Gamification in the classroom: Examining the impact of gamified quizzes on student learning," Computers & Education, vol. 144, Article ID 103666. 2020. https://doi.org/10.1016/j.compedu.2019.103666

[29] B. Sibbald and C. Roberts, "Understanding controlled trials: Crossover trials," BMJ, vol. 316, pp. 1719, 1998. https://doi.org/10.1136/bmj.316.7146.1719

[30] N. G. Reich and A. M. Milstone, "Improving efficiency in cluster-randomized study design and implementation: Taking advantage of a crossover," Open Access Journal of Clinical Trials, vol. 2014, no. 6, pp. 11–15, 2014. https://doi.org/10.2147/OAJCT.S56730

[31] P. M. Lightbown and N. Spada, How Languages are Learned (4th Ed.) Oxford, England: Oxford University Press, 2013.

[32] T. M. Lim and M. M. Yunus, "Teachers' perception towards the use of Quizizz in the teaching and learning of English: A systematic review," Sustainability 2021, vol. 13, no. 11, p. 6436, 2021. https://doi.org/10.3390/su13116436

[33] A. Crowe, C. Dirks, and M. Wenderoth, "Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology." CBE Life Science Education, vol. 7, pp. 368–381, 2008. https://doi.org/10.1187/cbe.08-05-0024

[34] A. Prunuske, L. Henn, A. Brearley, and J. Prunuske, "A randomized crossover design to assess learning impact and student preference for active and passive online learning modules," Medical Science Educator, vol. 26, pp. 135–141, 2016. https://doi.org/10.1007/s40670-015-0224-5

[35] A. Nehring, K. H. Nowak, A. U. Belzen, and R. Tiemann, "Predicting students' skills in the context of scientific inquiry with cognitive, motivational, and sociodemographic variables," International Journal of Science Education, vol. 37, no. 9, pp. 1343–1363, 2015. https://doi.org/10.1080/09500693.2015.1035358

[36] H. Alomari and R. S. Ekhlawi, "Reliability estimates of composites achievement tests in mathematics using the stratified alpha formula," European Journal of Scientific Research, vol. 155, no. 4, pp. 409–421, 2020. https://www.europeanjournalofscientificresearch.com/issues/PDF/EJSR_155_4_05.pdf

[37] S. L. Bretz and K. J. Linenberger, "Development of the enzyme-substrate interactions concept inventory," Biochemistry and Molecular Biology Education, vol. 40, no. 4, pp. 229–233, 2012. https://doi.org/10.1002/bmb.20622

[38] L. M. McClary and S. L. Bretz, "Development and assessment of a diagnostic tool to identify organic chemistry students' alternative conceptions related to acid strength," International Journal of Science Education, vol. 34, no. 15, pp. 2317–2341, 2012. https://doi.org/10.1080/09500693.2012.684433

[39] D. Wren and J. Barbara, "Psychometric analysis of the thermochemistry concept inventory," Chemistry Education Research and Practice, vol. 15, no. 3, p. 380–390, 2014. https://doi.org/10.1039/C3RP00170A

[40] J. Reeve and C. M. Tseng, "Agency as a fourth aspect of students' engagement during learning activities," Contemporary Educational Psychology, vol. 36, no. 4, pp. 257–267, 2011. https://doi.org/10.1016/j.cedpsych.2011.05.002

[41] A. Montenegro, "Understanding the concept of student agentic engagement for learning," Colombian Applied Linguistics Journal, vol. 19, no. 1, pp. 117–128, 2017. https://doi.org/10.14483/calj.v19n1.10472

[42] S. Senn, "The AB/BA crossover: past, present and future?," Statistical Methods in Medical Research, vol. 1994, no. 3, pp. 303–324, 1994. https://doi.org/10.1177/096228029400300402

7 Author

Kyung-Mi O is a Professor in the Department of English at Dongduk Women's University, Seoul, Korea. She received a doctoral degree in language, literacy, and technology specializing in language testing from Teachers College, Columbia University. Her major research interests have been language testing, second language acquisition, and language assessment and technology. Her current research project focuses on technology-enhanced formative assessment. Email: kmo@dongduk.ac.kr

Article submitted 2022-07-22. Resubmitted 2023-03-29. Final acceptance 2023-04-13. Final version published as submitted by the authors.