Huang, B., & Hew, K. F. (2021). Using Gamification to Design Courses: Lessons Learned in a Three-year Design-based Study. *Educational Technology & Society*, *24* (1), 44–63.

Using Gamification to Design Courses: Lessons Learned in a Three-year Design-based Study

Biyun Huang[1*] and Khe Foon Hew[2]
[1]Centre for Learning Sciences and Technologies, The Chinese University of Hong Kong, Hong Kong SAR, China // [2]Faculty of Education, The University of Hong Kong, Hong Kong SAR, China // lucyhuang99@cuhk.edu.hk // kfhew@hku.hk
[*]Corresponding author

(Submitted August 21, 2020; Revised October 15, 2020; Accepted December 5, 2020)

ABSTRACT: A design-based study was conducted in iterative cycles to test the effectiveness of the updated goal-access-feedback-challenge-collaboration (updated-GAFCC) gamification design model. The test-bed was a 10-week undergraduate introductory information management course. Students from three consecutive school years participated in the study, with the control group studying the conventional course without gamification (first year), treatment group_1 studying a gamified course following the original GAFCC model (second year), and treatment group_2 studying an optimized gamified course following the updated-GAFCC model (third year). The results of the design-based study indicated that (i) the updated-GAFCC model and the GAFCC model were effective in enhancing students' learning achievements and task completion; (ii) the updated-GAFCC model was more successful in generating higher quality thinking artifacts than the GAFCC model; (iii) there were fewer lower-quality submissions in the updated-GAFCC condition than in the GAFCC condition; and (iv) 89% of the interviewed students in the updated-GAFCC condition were satisfied with the overall learning design, and felt that the gamified learning activities facilitated their learning. Overall, the findings contribute to our understanding of how pedagogical strategies can be incorporated into the theory-based design model to optimize learning experiences and academic outcomes.

Keywords: Gamification, Design model, Design-based research, Learning performance, Long-term

1. Introduction

Gamification is usually defined as the use of game-like elements, such as badges, leaderboards, and points, in a non-game context (Deterding et al., 2011). Many researchers and practitioners have described gamification as a promising means of motivating students and optimizing their learning outcomes (Bai et al., 2020; Zainuddin et al., 2020). Empirical gamification literature covers K-12 education (e.g., Jong, 2019), university-level education (e.g., Baydas & Cicek, 2019), and short-term online training (e.g., Li et al., 2012). This line of research has investigated the role that gamification plays in behaviorally engaging participants (e.g., Ding, 2019; Barata et al., 2017), improving their academic performance (e.g., Jo et al., 2018), promoting their affective engagement (e.g., Sailer & Sailer, 2020), and catering for learners with different characteristics (e.g., Lopez & Tucker, 2019). In addition, some studies have examined the correlation between the setting of goals and performance in tasks (e.g., Landers et al., 2017).

Overall, the findings of the empirical literature were not consistent, with gamification having been found to be effective in certain situations and ineffective in others. Yildirim (2017) incorporated gamification into a teaching methods course, and reported that students in the gamified course showed higher improvements in learning than the control group. Lo and Hew (2018) found the submission rate of optional assignments in a gamified-flipped mathematics course was higher than that of a traditional course. However, Kyewski and Kraemer (2018) found that the students in the badge conditions did not participate more than the students in the no-badge condition. The inconsistent results suggest that gamification itself does not necessarily improve learning, and that the learning outcomes depend on how gamification is designed and implemented (Dichev & Dicheva, 2017). Thus, we need to explore methods that can optimize the use of gamification in education.

Although design-based research follows an iterative process in systematically examining and refining innovations, and thus produces design principles that can guide similar research efforts (Amiel & Reeves, 2008), this approach is often overlooked in the field of gamification. To date, few studies have used design-based research methods to explore the impact of gamification strategies, especially over a long period (Zainuddin et al., 2020). Moreover, the few empirical design-based studies (e.g., Lee et al., 2013; Cai et al., 2016) reported only the results of the first iterative cycles, not the second cycles, as is conventionally expected. To the best of our knowledge, no design-based studies have iteratively evaluated and refined a theory-based design model that can be used to guide other research studies. To address this gap, our study aims to test and refine a theory-based goal-access-feedback-challenge-collaboration (GAFCC) model using a three-year design-based research approach.

2. Design-based research in gamification

Design-based research seeks to solve real classroom problems that involve collaboration between researchers and practitioners. Design-based studies are typically conducted in iterative cycles of (1) analysis and exploration, (2) design and construction, and (3) evaluation and reflection (Figure 1; McKenney & Reeves, 2012). As shown in Figure 2, these processes facilitate the development of interventions that mature with each iteration. This approach allows the gamification interventions to develop over a longer time and can thus enhance our understanding of the interventions.

Lee et al. (2013) were among the first researchers to use a design-based approach to evaluate the effectiveness of a gamified product. Their study focused on the use of challenges, missions, points, badges, avatars, and leaderboards in the online social system Greenify, which was used to educate adults about issues relating to climate change. The results of their pilot study showed that the participants considered the system to be motivational and fun (79.3%). The system was also found to encourage users to create meaningful, and relevant content for their peers, and increase their awareness of how their behavior and lifestyle might affect the environment (46.2%). The pilot results were quite positive, but no further iterative cycles were conducted.

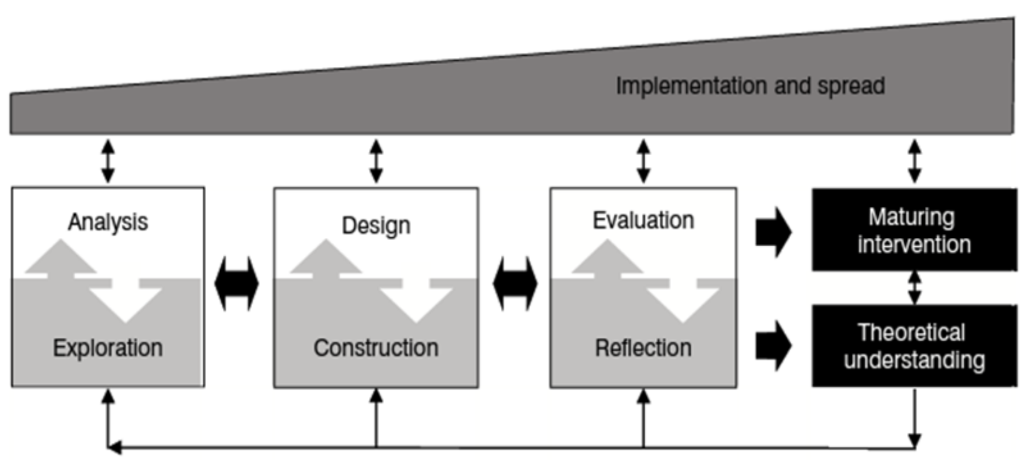


Figure 1. Generic model for design-based research in education (McKenney and Reeves, 2012)

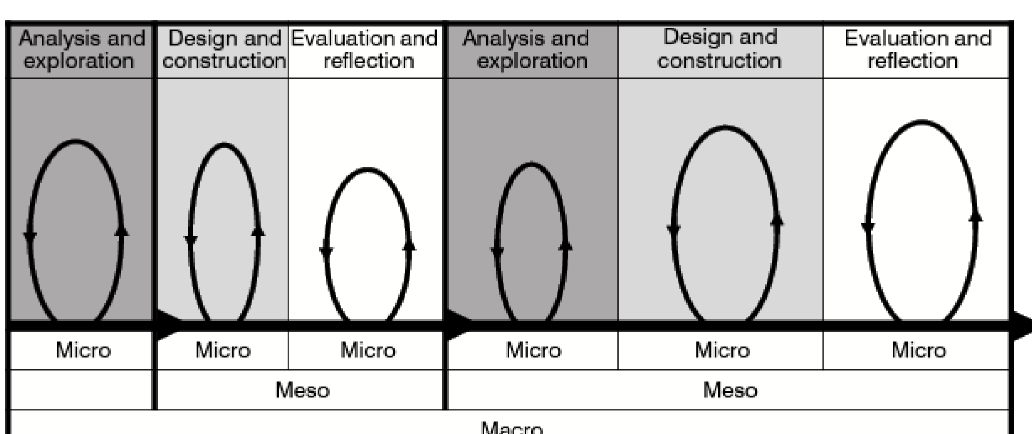


Figure 2. Micro-, meso, and macro-cycles in design-based research (McKenney and Reeves, 2012)

Barata et al. (2017) gamified a multimedia production course and varied their gamification strategies in three iterative cycles. In the first year, they used experience points, leaderboards, badges, and levels. The students were marked on their gamified activities, with the total score contributing to 10% of their final grades. In the second year, the researchers increased the weight of the gamified activities to 20% and added several gamification features, such as the optional Skill Tree and Talkative activities. In the third year, the researchers removed certain quizzes and other challenges to reduce the workload. The results for the third year revealed that the students were more engaged with the activities, and that most students found the course more interesting than other courses. However, the third year students had lower grades (75.4%, *SD* = 14.4) than the first year (79.5%, *SD* = 9.3) and second year students (88.3%, *SD* = 8.1). The results indicated that autonomy and choice, suitable levels of difficulty, and team-based challenges positively enhanced students' participation and satisfaction but did not enhance their learning performance.

Aguilar et al. (2018) iteratively tested a gamification design in a political theory course. In the first round, they allowed the students to choose their assignments, which comprised 60% of the final grade, and power-ups (i.e., awards that students could earn to either increase their grades, make up for absences, or unlock assigned tasks). The first cycle results indicated that the students' perceived autonomy was influenced by the perceived fairness of the grading system and the difficulty of the tasks. The results of the second-round investigation confirmed the initial results. The two rounds showed that gamification supported autonomy and promoted deeper engagement. Overall, the study provided evidence that autonomy is an important element of gamification design.

In summary, although design-based research has helped identify factors (e.g., autonomy) that can improve the motivation of students, none of these studies have contributed to the development and testing of a theory-driven gamification design model. Our study fills this research gap by validating a theory-based meta-model of gamification that will pave the way for future gamification designs.

3. Theoretical framing

This study uses a design-based research framework to evaluate the effectiveness of a meta-model of gamification. The theoretical framework guiding this study is the GAFCC model originally proposed by Huang and Hew (2018). The meta-model consists of two integral parts. The first part introduces the theoretical underpinnings of gamification design and explains how game mechanics can be used to construct a motivational learning environment. The GAFCC model is derived from five motivation theories: goal setting theory (Locke & Latham, 2002), flow theory (Csikszentmihalyi, 1990), self-determination theory (Deci & Ryan, 2000), social comparison theory (Festinger, 1954), and behavioral reinforcement theory (Skinner, 1953). The model highlights that *goal* setting, access to choices and the development of competence, *feedback* on performance and status, (suitable levels of) *challenge*, and a sense of connectedness developed during *collaboration* are important motivational needs. Moreover, specific game mechanics (e.g., badges, leaderboards) corresponding to individual motivational needs can be integrated into learning environments to engage students and encourage expected behaviors (Figure 3).

The second part of the GAFCC model uses the five-step design procedure to gamify a course. This part emphasizes the steps for building alignment between motivational needs, instructional objectives, learning activities, and gamification strategies (Figure 4). We decided to use the GAFCC model as the framework for our design-based study because the model is well grounded in the five motivation theories, and interprets people's motivational needs from the individual and community perspectives. The five-step design procedure is also easy to follow and useful in guiding the gamifying process.

The test-bed was a 10-week undergraduate introductory information management course. The first cycle (pilot study 1) spanned over two years. Students in the first year participated in a non-gamified course. Students in the second year participated in a gamified course based on the GAFCC design model (i.e., treatment group_1). In the second cycle (main study 2), students in the third year participated in a gamified course following the updated-GAFCC design model (i.e., treatment group_2; Figure 5). Pilot study 1 has been reported elsewhere (Huang et al., 2019). In this study, we concentrate on main study 2 and compare the implementation results for treatment group_2 with the results for the control group and treatment group_1. Before describing study 2 in detail, we first summarize the main findings and concerns in pilot study 1.

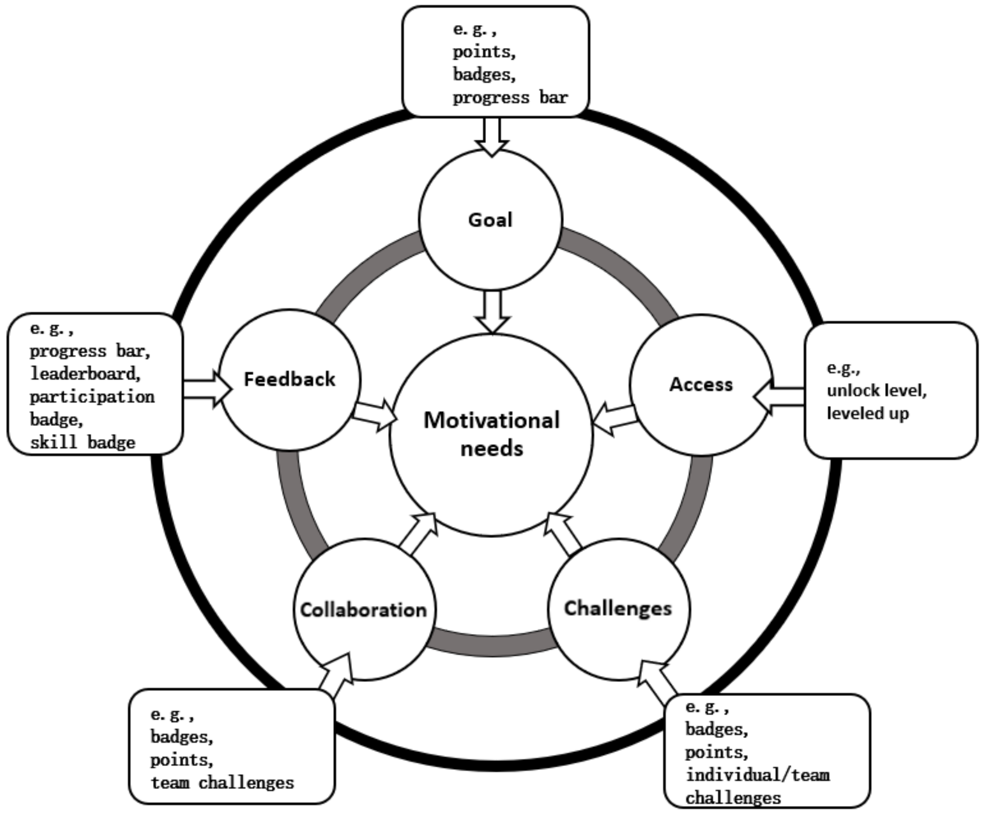


Figure 3. GAFCC model (Huang and Hew, 2018)

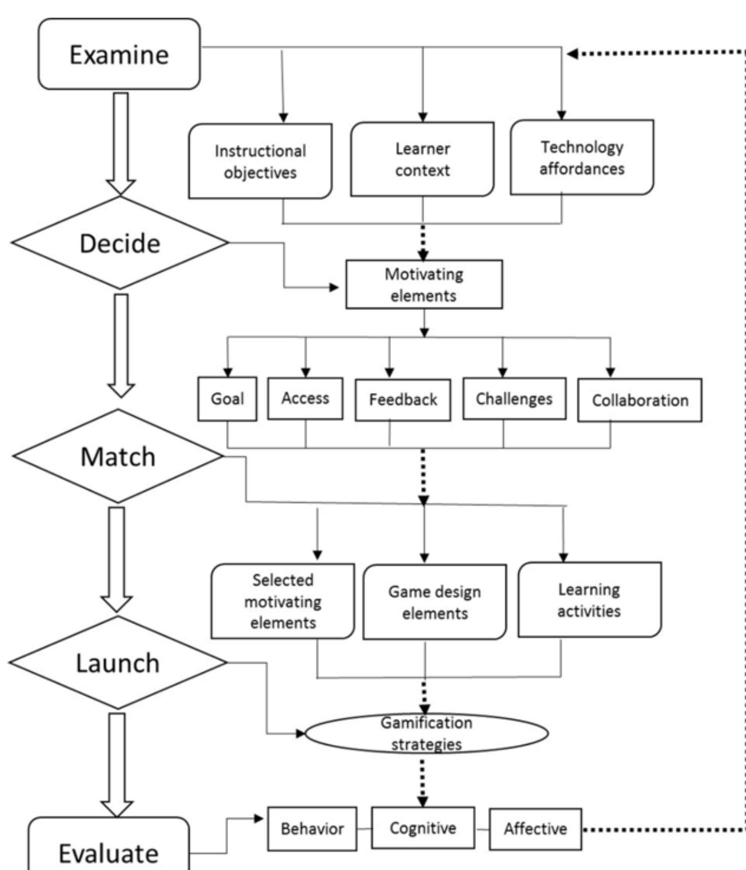


Figure 4. Five-step gamification design procedure (Huang and Hew, 2018)

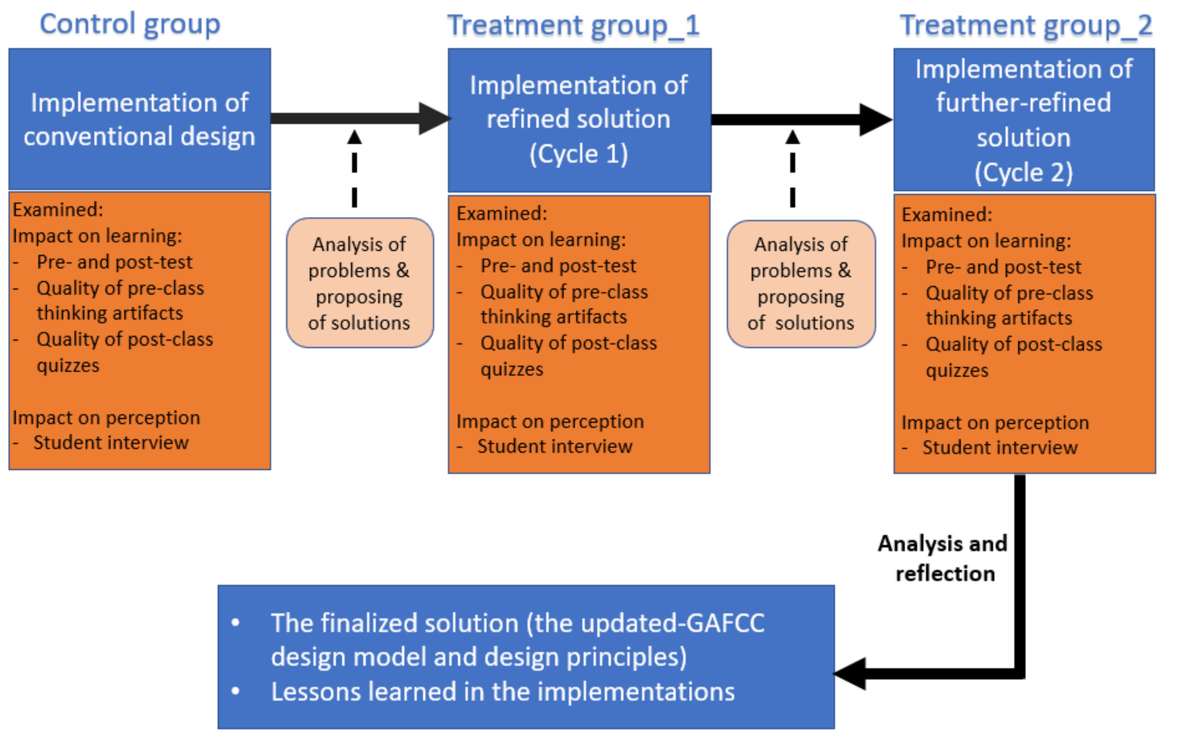


Figure 5. Design of the design-based research project

4. Pilot Study 1: Examining the effects of gamification on student learning and participation

Study 1 was an exploratory pilot study that sought to determine the effects of gamification (if any) on student learning and participation. Forty-eight undergraduates attended an information management course without gamification (control group), while another 48 attended the same course with gamification (treatment group_1). The GAFCC model was used to design treatment group_1 based on the use of badges, levels, and a leaderboard, as shown in Figure 6. Figure 7 shows the different levels used in treatment group_1.

Two types of badges (participation badges and quality badges) were used in treatment group _1. Participation-based badges (e.g., early bird, super-efficient) were used to motivate the students to complete more activities. Quality-based badges (e.g., movie, coffee, tour package, champion) were used to encourage the students to relate their contributions to the contents learned in the course, and provide additional insights on the course-related topics. None of the quality badges could be exchanged for tangible products. The number of badges each student collected was not linked to the course credits.

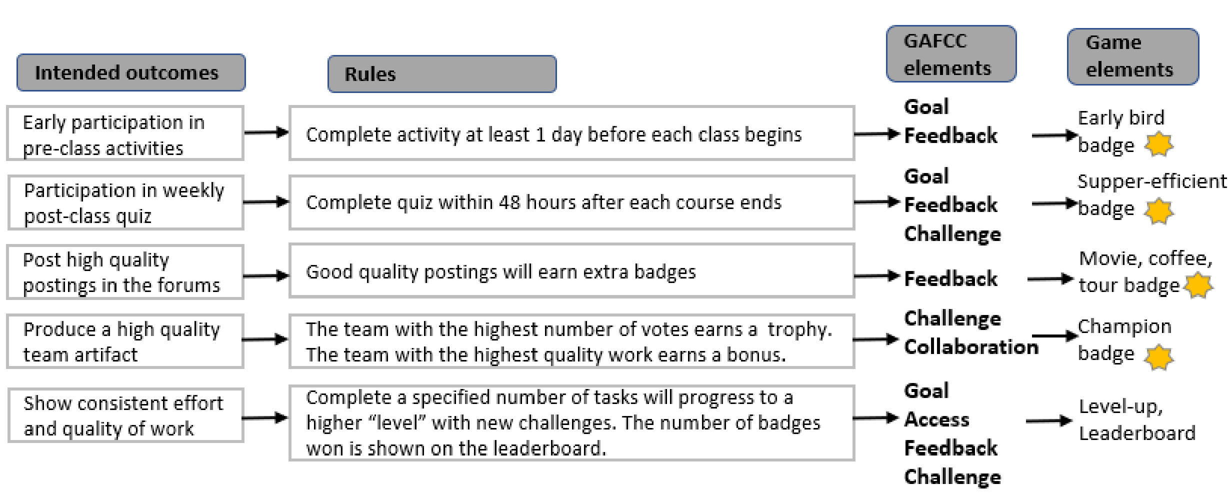


Figure 6. Intended outcomes, rules, GAFCC elements, and game elements in treatment group_1

The results suggested that the GAFCC model was effective in improving student participation and learning. More students in the GAFCC group completed the pre- and post-class activities than students in the control group. In addition, treatment group_1 submitted significantly more in-depth level pre-class thinking artifacts than the control group. Thus, the effectiveness of the positive outcome is, in all likelihood, associated with the fulfillment of the core psychological needs of the students.

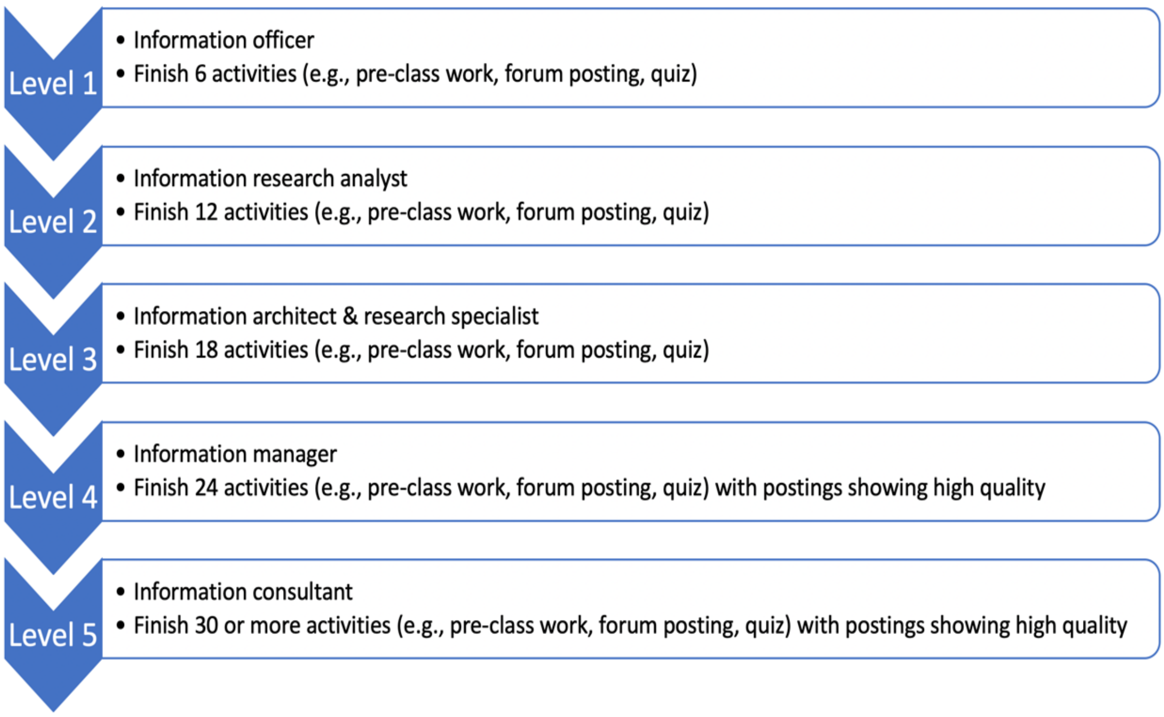


Figure 7. Game mechanics (levels) used in treatment group_1

However, a main concern was that some students frequently submitted low quality postings solely for the purpose of attaining participation badges. A number of students also perceived that the criteria for winning quality-based badges were not explicit enough. Although written explanations of how to win quality-based badges were provided in a slide on the course forum, few students clicked the slide and viewed the explanations. Based on the findings of pilot study 1, we determined that although gamification can help students set goals and provide feedback, the game mechanics may not sufficiently promote learning because students can trick the system by submitting low-quality content.

5. Main Study 2: Optimizing the use of gamification

5.1. Method

In response to the findings and concerns in pilot study 1, we proposed an updated-GAFCC model. We assumed that gamification design could be supplemented by other pedagogical strategies, such as teacher feedback, to optimize students' quality of submission. Thus, in the updated-GAFCC model, the five-step design process is supplemented with pedagogical strategies (Figure 8). We tested the effectiveness of the updated-GAFCC model with a new cohort (i.e., treatment group_2) and compared the results with those for treatment group_1 and the control group, as shown in Figure 9. Before beginning the experiment, we briefed the students about the research project, and invited them to sign a consent form. In total, 50 students participated in treatment group_2. Pre-test results showed that there were no significant differences in the students' prior knowledge between the three groups.

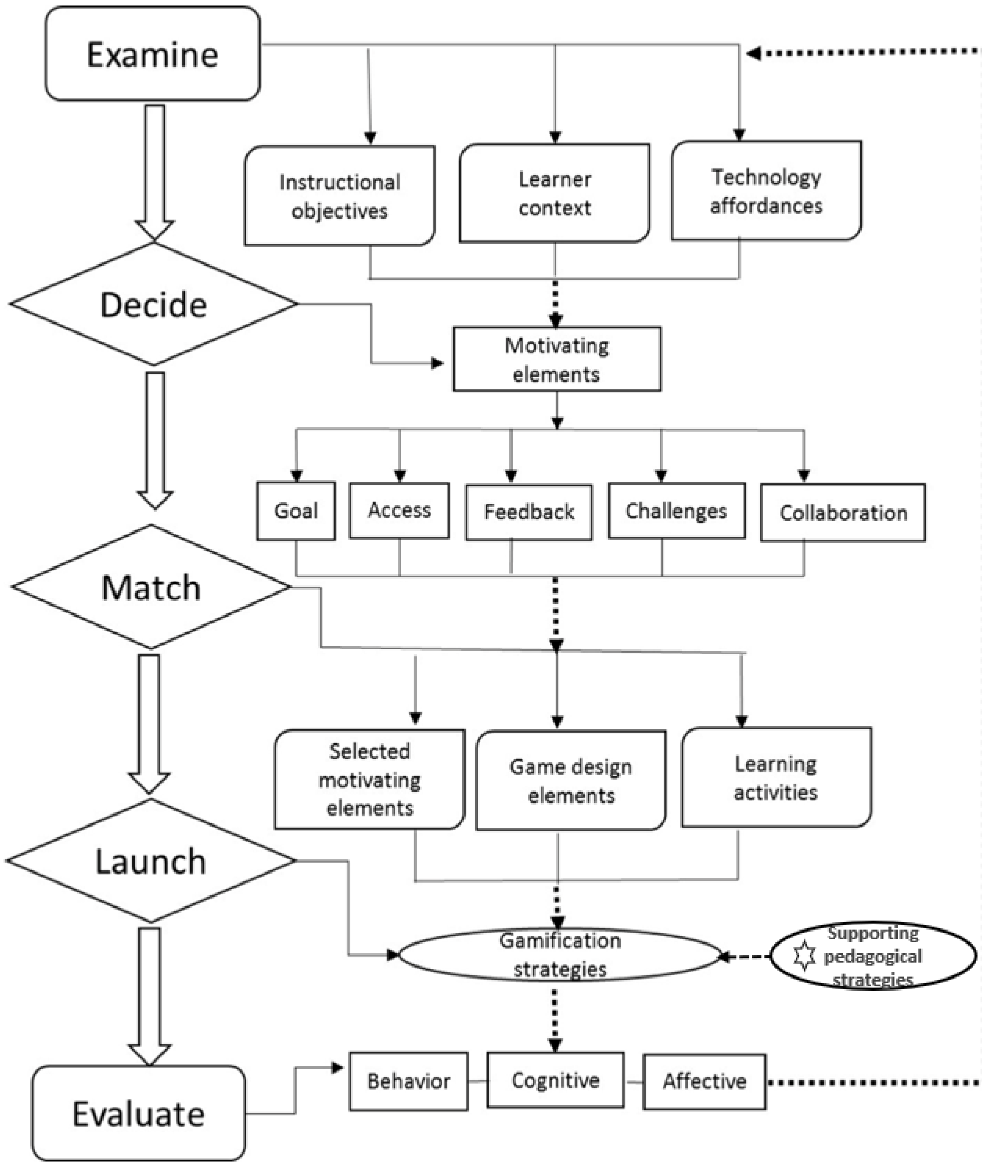
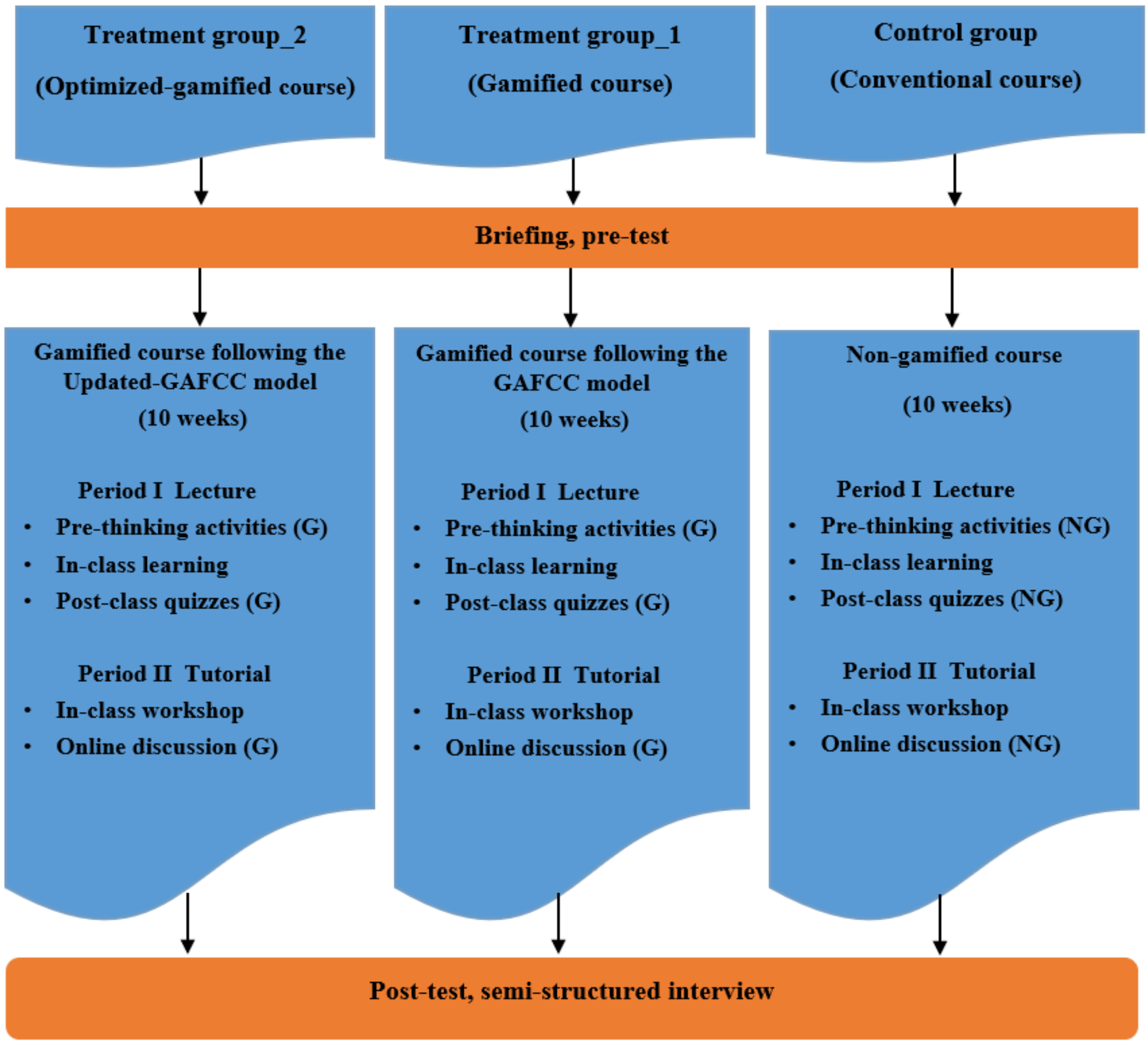


Figure 8. The five-step design process in the updated-GAFCC model



Note. "G" refers to "gamified," "NG" refers to "non-gamified."
Figure 9. Procedures of the design-based research

5.2. Instructional design of Study 2

The design process consisted of the following phases.

5.2.1. Phase 1: Analysis of the problems revealed in Study 1

In the analysis phase, the researcher collaborated with the course instructor to solve the problems that arose in Study 1. Although most of the students in Study 1 preferred to learn in a gamified environment, they reported a number of specific problems (Figure 10). One student stated that the timing of the vote for the best team was too late, and suggested holding it earlier. In addition to the problems raised by the students, the researcher and the course instructor were aware that insufficient feedback on student work was provided in Study 1, and the leaderboard display was not convenient to access.

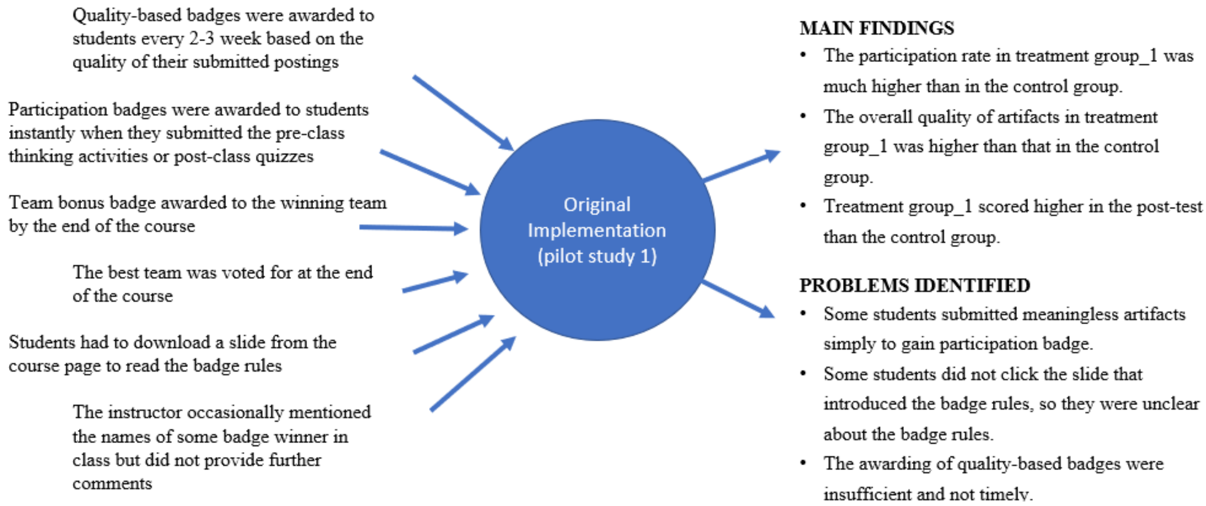


Figure 10. Summary of the implementation, findings, and problems in pilot study 1

5.2.2. Phase 2: Design and construction of the further-refined solutions in practice

The same intended outcomes and game elements (Figure 6) were used in the main study. The researchers reflected on the problems identified in the previous cycle, and proposed to refine the design by following the updated-GAFCC design model. With respect to the gamification strategies, the researchers decided to do the following:

- Provide a more direct way to visualize the badge rules and different levels. The badge rules and levels were displayed directly on the main announcement page, which could be immediately seen upon login, so the students did not need to download a slide to view the rules or levels (Figure 11).

- Assign badges and team bonuses more frequently. Quality-based badges were assigned to students on a weekly basis, rather than every two to three weeks, and team bonuses were assigned on a weekly basis rather than at the end of the course.

- Move voting for the best team to the middle of the course rather than at the end of the course.

With respect to the pedagogical strategies, it was decided that the instructor would provide elaborative feedback to students at the beginning of the class at least once every two weeks. Feedback is most beneficial when it guides the modification of thinking and provides students with opportunities for refinement (National Research Council [NRC], 2000). Although gamification can provide evaluative feedback regarding what learners have achieved, such as granular feedback (e.g., displaying points linked to each positive action), sustained feedback (e.g., using a performance graph to display progress), and cumulative feedback (e.g., tracking and assessing a series of learner actions and presenting the results on a leaderboard) (Rigby & Ryan, 2011), it does not convey complex information on what the learners can do to improve their thinking. For this reason, supplementing the gamified feedback with elaborative feedback from the teacher would likely help the students to better understand how well they have performed and how they can improve the quality of their submissions. For example, in the pre-class thinking activities, the students had made their thinking visible to themselves, teachers, and peers. Thus, if the students subsequently listened to the teacher's elaborative feedback, they would likely to reflect on their thinking and refine their submissions in future weeks (NRC, 2000). In the earlier study, the instructor only occasionally mentioned the names of the students who had won specific badges and did not make any further comments. In this cycle, the instructor provided examples of high-quality content based on the students' submissions and explained why the examples were considered to be of high quality. See Figure 12. for the procedure of the elaborative feedback.

5.2.3. Phase 3 Test of the proposed solutions

After updating the design, the course instructor implemented the refined gamification strategies and pedagogical strategies in treatment group_2. The efficacy of the design was evaluated by analyzing the students' learning artifacts, participation data, and interview results.

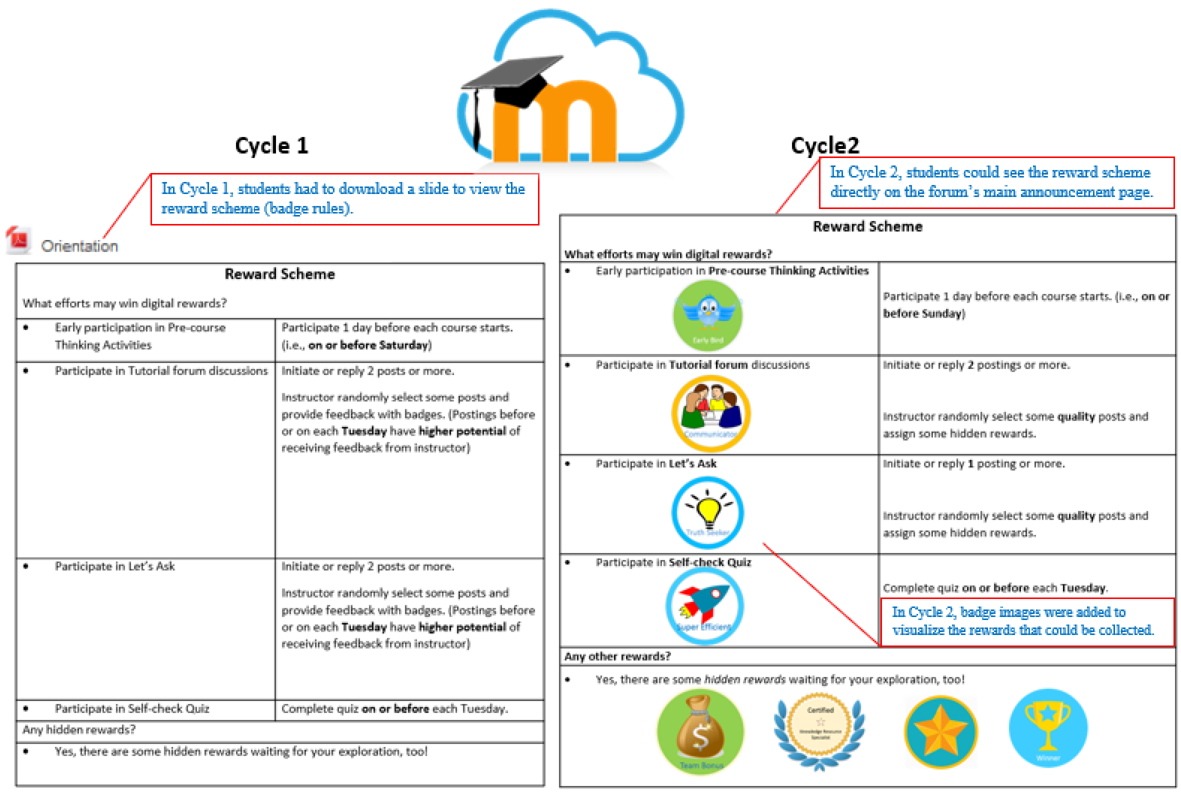


Figure 11. Display of badge rules on Moodle in Cycle 1 and Cycle 2

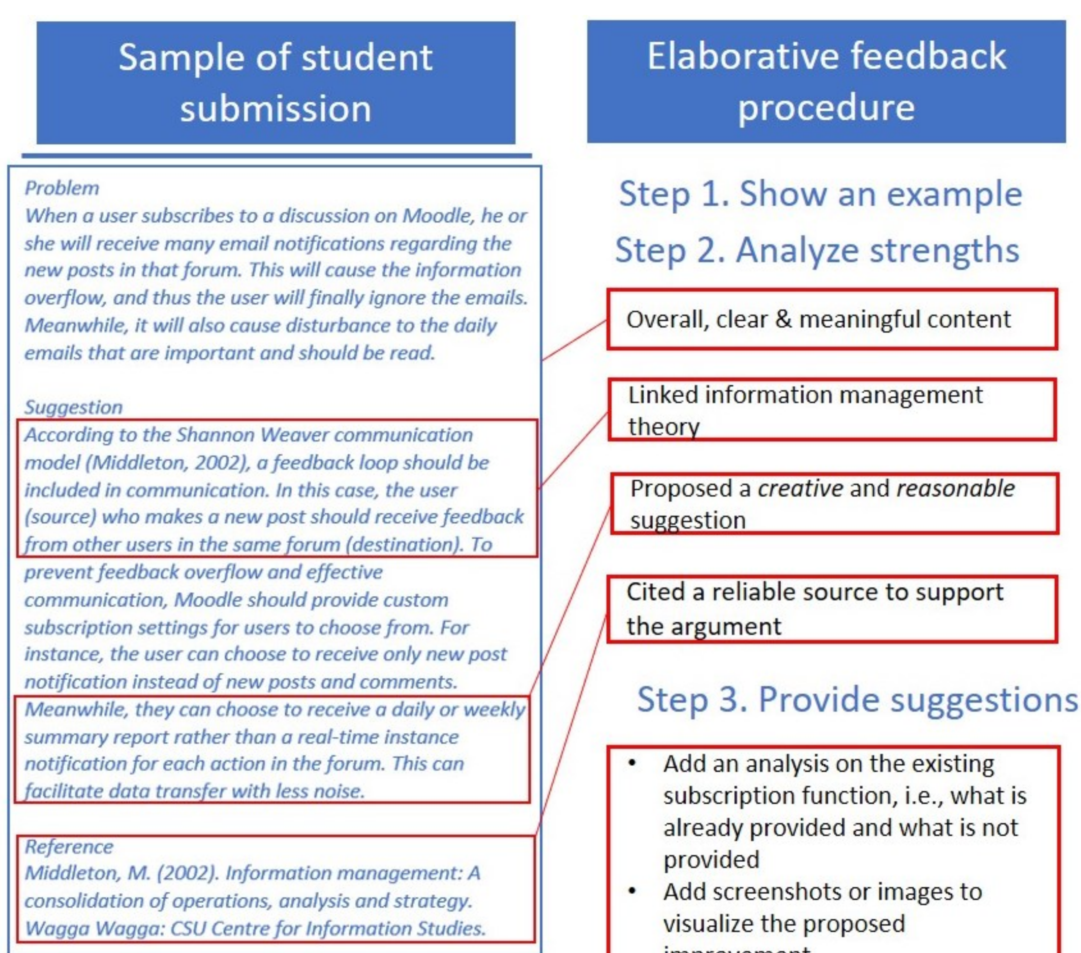


Figure 12. A sample of submissions and the procedure of elaborative feedback in Cycle 2

6. Results

To compare the effectiveness of the conventional approach, the GAFCC model, and the updated-GAFCC model, we evaluated the students' participation data and the quality of the artifacts. The participation indicators were the completion of pre-class thinking activities and post-class quizzes. The quality indicators were the pre- and post-test scores, levels of critical thinking exhibited in the thinking activities, and scores for the post-class quizzes.

6.1. Impact on participation

6.1.1. Pre-class activity completion

Pre-class thinking activities were assigned to the students each week, except for week 1 and week 10. In weeks 4 and 5, only one pre-thinking activity (i.e., pre-thinking activity 4-5) was assigned. A technical problem occurred in pre-thinking activity 4-5 in the first year, and the students could not submit their work for the activity. Hence, pre-thinking activity 4-5 was not included in the analysis. Chi-square tests of independence showed that there were significant differences in the completion rates of the three groups during four of the six weeks examined, i.e., weeks 6, 7, 8, and 9. Treatment group_2 and treatment group_1 were more likely to complete the activities before the due date than the control group. For example, a significant result was found in week 6 ($X^2(2) = 33.97$, $p < .001$), which showed that treatment group_2 (76%) and treatment group_1 (75%) were more likely to complete the activities before the due date than the control group (25%) (Figure 13 and Table 1).

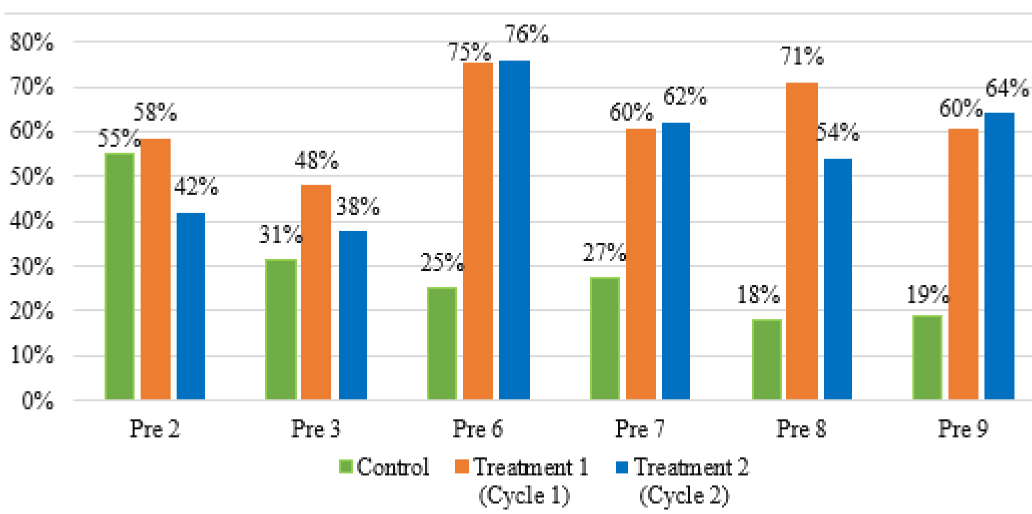


Figure 13. Pre-class thinking activities completed before the weekly due dates (Cycles 1 & 2)

Table 1. Completion of the pre-class thinking activities before the weekly due dates (Cycles 1 & 2)

Activity	Condition	Completed before the due date	Chi-square
Pre_think 2	Treatment2	42%	$X^2(2) = 2.84, p > .05$
	Treatment1	58%	
	Control	55%	
Pre_think 3	Treatment2	38%	$X^2(2) = 2.84, p > .05$
	Treatment1	48%	
	Control	31%	
Pre_think 6	Treatment2	76%	$X^2(2) = 33.97, p < .001$
	Treatment1	75%	
	Control	25%	
Pre_think 7	Treatment2	62%	$X^2(2) = 15.05, p < .05$
	Treatment1	60%	
	Control	28%	
Pre_think 8	Treatment2	54%	$X^2(2) = 27.20, p < .001$
	Treatment1	71%	
	Control	18%	
Pre_think 9	Treatment2	64%	$X^2(2) = 24.55, p < .001$
	Treatment1	60%	
	Control	19%	

Note. Treatment2 (n = 50); Treatment1 (n = 48); Control (n = 48).

6.1.2. Post-class quiz completion

Post-class quizzes were provided to the students each week, except for weeks 1 and 10. In weeks 4 and 5, only one quiz (i.e., quiz 4-5) was assigned to the students. Chi-square tests of independence showed significant differences in the completion rates of treatment group_2, treatment group_1, and the control group in five of the seven weeks, namely weeks 4-5, 6, 7, 8, and 9. Treatment group_2 and treatment group_1 were more likely to complete the activity before the due date than the control group. For example, a significant interaction was found in week 4-5 ($X^2(2) = 8.41$, $p < .05$), which showed that treatment group_2 (32%) and treatment group_1(25%) were more likely to complete the activity before the due date than the control group (15%) (see Figure 14 and Table 2).

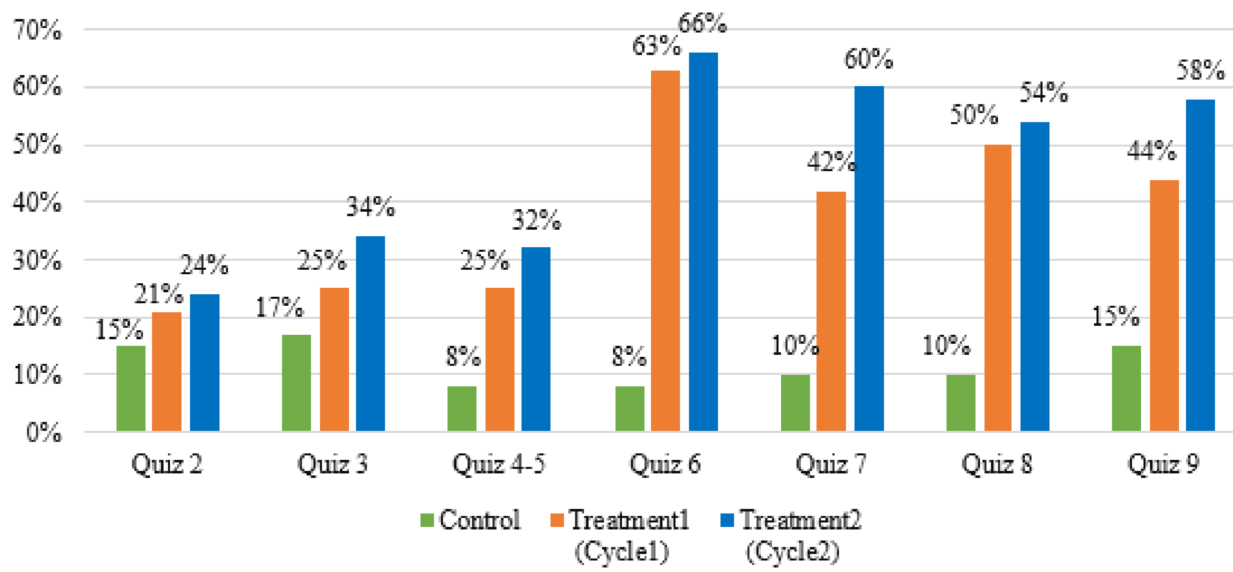


Figure 14. Post-class quizzes completed before the weekly deadlines (Cycles 1 & 2)

Table 2. Completion of post-class quizzes before the weekly due dates (Cycles 1 & 2)

Post-class quiz	Condition	Completed before the deadline	Chi-square
Quiz 2	Treatment2	24%	$X^2(2) = 1.41$, $p >.05$
	Treatment1	21%	
	Control	15%	
Quiz 3	Treatment2	34%	$X^2(2) = 3.89$, $p >.05$
	Treatment1	25%	
	Control	17%	
Quiz 4-5	Treatment2	32%	$X^2(2) = 8.41$, $p <.05$
	Treatment1	25%	
	Control	8%	
Quiz 6	Treatment2	66%	$X^2(2) = 40.74$, $p < .001$
	Treatment1	63%	
	Control	8%	
Quiz 7	Treatment2	60%	$X^2(2) = 26.13$, $p < .001$
	Treatment1	42%	
	Control	10%	
Quiz 8	Treatment2	54%	$X^2(2) = 23.78$, $p < .001$
	Treatment1	50%	
	Control	10%	
Quiz 9	Treatment2	58%	$X^2(2) = 20.06$, $p < .001$
	Treatment1	44%	
	Control	15%	

Note. Treatment2 ($n = 50$); Treatment1 ($n = 48$); Control ($n = 48$).

6.2. Impact on participation

6.2.1. Pre-test and post-test scores

As a normality test indicated that the pre-test scores were not evenly distributed, a nonparametric Kruskal-Wallis H test was used to examine the differences between the three groups. The Kruskal-Wallis H test revealed that there were no statistically significant differences in the pre-test scores of the different conditions, $X^2(2) = 2.27$, $p > .05$, with a mean rank score of 56.85 for treatment group_2 ($n = 40$), 65.21 for treatment group_1 ($n = 39$), and 58.08 for the control group ($n = 40$).

The post-test scores of the three groups were also evaluated. Kruskal-Wallis H test was used to examine the differences between the groups. The results showed that there were statistically significant differences in the post-test scores between the different conditions, $X^2(2) = 14.52$, $p < .001$, with a mean rank score of 56.66 for treatment group_2 ($n = 35$), 60.09 for treatment group_1 ($n = 38$), and 36.32 for the control group ($n = 30$). A post-hoc Bonferroni test showed there were significant differences between treatment group_2 ($M = 4.3$, $SD = 1.26$) and the control group ($M = 3.60$, $SD = 1.22$), $p = .008$. Similarly, there were significant differences between treatment group_1 ($M = 4.43$, $SD = 1.10$) and the control group ($M = 3.60$, $SD = 1.22$), $p = .001$. Both treatment group_2 and treatment group_1 outperformed the control group in the post-test.

6.2.2. Levels of critical thinking in pre-class thinking activities

The quality of the pre-class thinking artifacts was evaluated based on the modified Taxonomy of Critical Thinking (Greenlaw & Deloach, 2003). Greenlaw and Deloach (2003) proposed that the level of critical thinking can be measured with respect to the degree of sophistication in an argument. We adapted the taxonomy to evaluate thinking artifacts (as shown in Figure 15) by retaining level 0 (off-the-topic); combining the original level 1(unilateral description) and 2 (simplistic argument) into one level as both deal with simple description or justification; retaining level 3 (explicit analysis); and merging levels 4 (theoretical inference) and 5 (empirical inference) into one level, namely theoretical and/or empirical inference. Level 6, merging value with analysis, was not included as it was beyond the scope of our investigation. However, we included creativity as an additional factor, because it was frequently observed in the artifacts of treatment group_2.

Level 0	**Level 1-Lower level**
❑**Off-the-topic, no submission, or submission after the due date**	❑**Mere repetition or simplistic arguments** *e.g., repeating question statements without adding new information or interpretation* *e.g., making confusing or ambiguous statements* *e.g., assertions without evidence, or giving an example to provide a simple explanation*
Level 2-Upper level	**Level 3-Uppermost level**
❑**Serious attempts to analyze an argument or list competing arguments with evidence** *e.g., serious argument, such as listing factors as evidence, exploring competing argument, citing anecdotal evidence, but with logical flaws* *e.g., serious argument with at least 1 creative perspective (which can shed light on the issue/or can bring inspiration to other students) but with logical flaws* *e.g., serious argument but not referring to theory or empirical data*	❑**Serious argument with a clear logical framework, strengthened by empirical evidence and/or theory** *e.g., use historical data, citing reliable resources, using theory to test the validity of an argument* *e.g., serious argument with two or more creative perspectives (which can shed light on the issue or bring inspiration to other students) and a clear logical framework*

Figure 15. Revised Taxonomy of Critical Thinking (adapted from Greenlaw and Deloach, 2003)

In the data analysis, a researcher first coded all of the submissions. An independent observer then randomly selected and coded 20% of the submissions. The Krippendorff's alpha test, a widely adopted test for examining the inter-rater reliability of content analysis results (Krippendorff, 2004), was conducted. The Krippendorff's alpha test result was $\alpha = 0.88$, indicating a substantial level of agreement between the two coders.

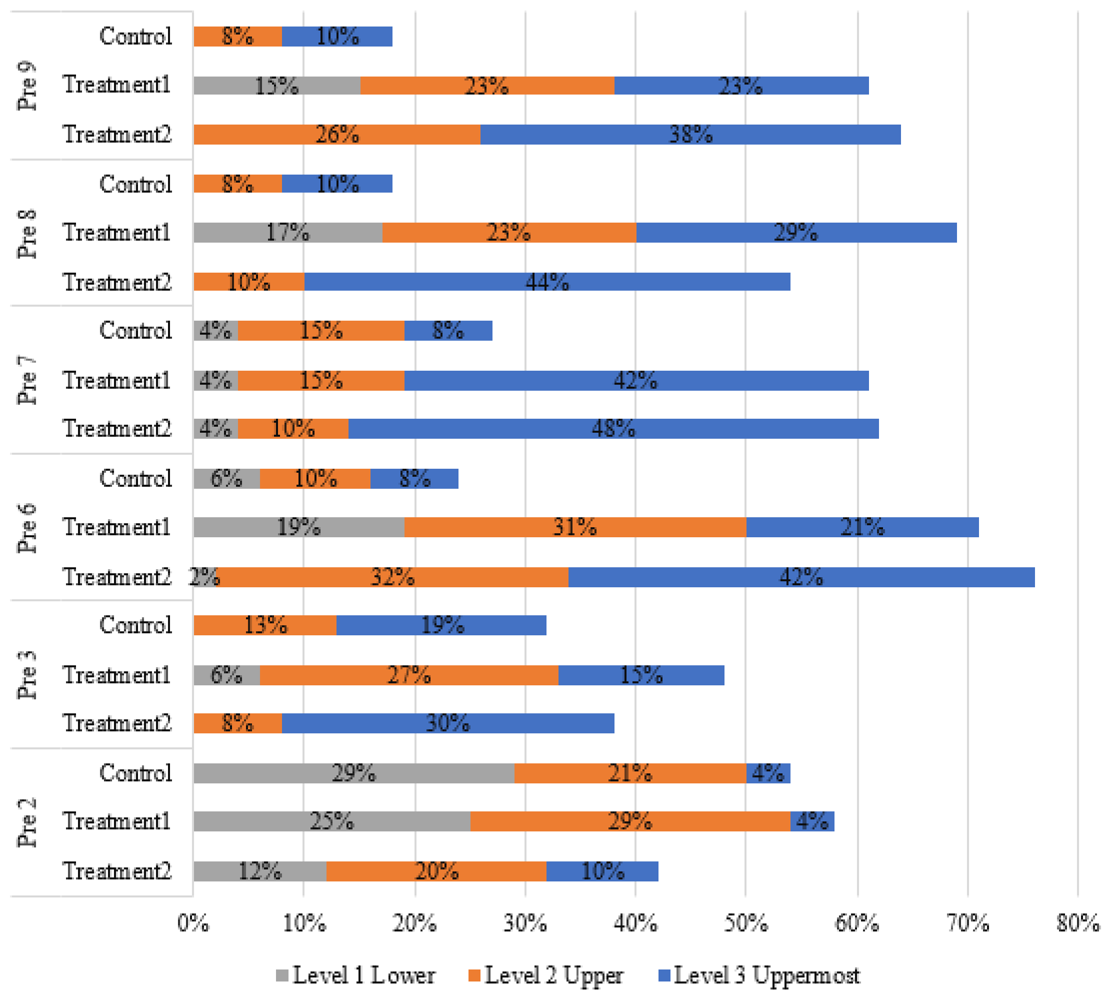


Figure 16. Distribution of the critical thinking levels in the pre-class thinking activities (Cycles 1 & 2)

Table 3. Quality of the pre-class thinking activities (Cycle 1 & 2)

Activity	Condition	Level 3 (uppermost)	Level 2 (upper)	Level 1 (lower)	Level 0 (low-late-no)	*p*-value
Pre_think 2	Treatment2	10%	20%	12%	58%	.246
	Treatment1	4%	29%	25%	42%	A, B = .134
	Control	4%	21%	29%	46%	A, C = .141
Pre_think 3	Treatment2	30%	8%	0%	62%	.019
	Treatment1	15%	27%	6%	52%	A, B = .008**
	Control	19%	13%	0%	69%	A, C = .432
Pre_think 6	Treatment2	42%	32%	2%	24%	< .001
	Treatment1	21%	31%	19%	29%	A, B = .013**
	Control	8%	10%	6%	75%	A, C < .001**
Pre_think 7	Treatment2	48%	10%	4%	38%	<.001
	Treatment1	42%	15%	4%	40%	A, B =.89
	Control	8%	15%	4%	73%	A, C < .001**
Pre_think 8	Treatment2	44%	10%	0%	46%	<.001
	Treatment1	29%	23%	17%	31%	A, B =.002**
	Control	10%	8%	0%	81%	A, C < .001**
Pre_think 9	Treatment2	38%	26%	0%	36%	<.001
	Treatment1	23%	23%	15%	40%	A, B = .020*
	Control	10%	8%	0%	81%	A, C < .001**

Note. *significant using $p < .05$; **significant using an adjusted-alpha of 0.017 (Bonferroni adjustment). Treatment2 (n = 50); Treatment1 (n = 48); Control (n = 48). A represents *treatment2*, B represents *treatment1*, and C represents *the control group*.

The descriptive data showed that treatment group_2 had more uppermost level submissions than treatment group_1 and the control group in all weeks. On average, 35% of treatment group_2, 22% of treatment group_1, and 10% of the control group students submitted uppermost level artifacts (Figure 16). As the expected counts of several cells were smaller than five, we conducted Fisher's exact tests rather than Chi-square tests of independence to compare the differences between the three groups since the Fisher's exact test is more appropriate for small values (Bolboacă et al., 2011). As shown in Table 3, the results indicated there were significant differences in the quality of the pre-class thinking artifacts of treatment group_2, treatment group_1, and the control group in five out of six weeks, namely weeks 3, 6, 7, 8, and 9. For example, in week 6, the students in treatment group_2 (42%) were more likely to submit uppermost level artifacts than those in treatment group_1 (21%) and the control group (8%), $p < .001$. Pairwise Fisher's exact tests showed that treatment group_2 submitted significantly more uppermost level artifacts than treatment group_1 in weeks 3, 6, 8, and 9, and than the control group in weeks 6, 7, 8, and 9.

6.2.3. Performance in post-class quizzes

The quality of the post-class quizzes completed before the due dates for treatment group_2 was compared with that for treatment group_1 and the control group. The quizzes submitted after the due dates and absent submissions were recorded as 0. Kruskal-Wallis H tests results indicated that there were significant differences in the post-class quiz scores of the three groups, except in weeks 2 and 3 (Table 4). For example, in week 4-5, the results showed that there were statistically significant differences in the post-class quiz scores of the three conditions, $X^2(2) = 9.67$, $p = .008$, with a mean rank score of 81.83 for treatment group_2 ($n = 50$), 75.38 for treatment group_1 ($n = 48$), and 62.95 for the control group ($n = 48$). Post-hoc Bonferroni tests showed that treatment group_2 outperformed the control group in weeks 4-5, 6, 8, and 9, and treatment group_1 outperformed the control group in weeks 6, 8, and 9. No significant differences were observed between treatment group_2 and treatment group_1, with the two groups performing equally well in the post-class quizzes.

Table 4. Post-class quiz scores completed before the due dates (Cycles 1 and 2)

	Group	*n*	Mean rank	Kruskal-Wallis test
Quiz 2	Treatment2	50	74.62	$X^2(2) = 0.848$, $p = .654$
	Treatment1	48	75.41	
	Control	48	70.43	
Quiz 3	Treatment2	50	80.5	$X^2(2) = 4.802$, $p = .091$
	Treatment1	48	73.38	
	Control	48	66.33	
Quiz 4-5	Treatment2	50	81.83	$X^2(2) = 9.67$, $p = .008$
	Treatment1	48	75.38	A, B = .88
	Control	48	62.95	A, C = .010**
Quiz 6	Treatment2	50	88.79	$X^2(2) = 36.15$, $p < .001$
	Treatment1	48	84.32	A, B = 1
	Control	48	46.75	A, C < .001**
Quiz 7	Treatment2	50	89.5	$X^2(2) = 24.72$, $p < .001$
	Treatment1	48	76.69	A, B = 0.24
	Control	48	53.65	A, C < .001**
Quiz 8	Treatment2	50	81.53	$X^2(2) = 20.10$, $p <.001$
	Treatment1	48	84.07	A, B = 1
	Control	48	54.56	A, C = .001**
Quiz 9	Treatment2	50	88.17	$X^2(2) = 20.83$, $p < .001$
	Treatment1	48	76.8	A, B = .37
	Control	48	54.92	A, C < .001**

Note. **significant using an adjusted-alpha of 0.017 (Bonferroni adjustment). Treatment2 ($n = 50$); Treatment1 ($n = 48$); Control ($n = 48$). A represents *treatment2*, B represents *treatment1*, and C represents *the control group*.

6.3. Students' perceptions of gamification

The students in treatment group_2 were invited to participate in semi-structured interviews after the completion of the course. Nine students (five females and four males) volunteered to participate in the interviews. Sample interview questions were "What is your general feeling about the course?" "What are the positive/negative aspects of gamification?" "Do you prefer to study in a gamified or non-gamified environment? Why?" "Do you have any suggestions for improving the gamification design?"

The interview results were coded by two researchers. The first researcher read through all the transcripts and determined the coding scheme, including the themes, initial codes, and coding procedures. Examples of each theme were provided to increase the consistency of classification. Then, the researcher coded the transcripts. The constant-comparative approach (Glaser, 1965) was adopted, and any new categories that emerged were given a new label. After receiving training, the second researcher randomly selected 20% of the transcripts and coded them. The percent agreement (Campbell et al., 2013) between the two raters was 83%. Discrepancies were discussed until agreements reached.

The responses to the question, "Do you prefer to study in a gamified or non-gamified environment" indicated that 89% (eight out of nine) of the students preferred to learn in a gamified learning environment. The remaining student stated that she would have preferred to learn in a gamified learning environment had the leaderboard been more competitive. As to the question of the positive aspects of gamification, students commented that gamification encouraged peer collaboration and interaction (100%, n = 9), provided recognition (100%, n = 9), added an element of fun to the course (78%, n = 7), and enhanced their competence (67%, n = 6).

The students explained how gamification motivated them to collaborate more. For example, one student commented:
I thought the forum activities were a good way for us to communicate with other students. Even though we did not meet face to face (after class), we could still communicate wherever and whenever. These badges brought us closer. Because we had to talk about tasks and communicate constantly. (S7)

The recognition conveyed through gamification stimulated the students to treat the activities seriously. One student stated:
Because it felt like we were winning a reward or extra marks. Moreover, some of the hidden rewards were sent by e-mail, and after I received them, I was surprised and felt that I had done a great job in the forum. This encouraged us to do more and take the tasks seriously. (S4)

The students also explained why it is important to include fun and enjoyment in a course. For instance, one student stated:
I enjoyed the gamified course because it was fun. I am a person who believes that if you enjoy learning, you will learn more than you would otherwise. People could enjoy this course, and I think that is essential. (S5)

In the updated-GAFCC condition, students focused more on making quality submissions rather than the quantity of their submissions (78%, n = 7). For example, one student expressed, "I put a lot more effort into this course than other courses. I spent a lot of time reading the notes so as to submit quality posts." (S1) Another student expressed that, "as gamification was used in this course, I decided that I not only had to use my own knowledge to reply on the forum, but also had to provide a quality response. It encouraged me to learn more about the topic. For example, I did more reading and things like that. It helped us discover different ways to improve our knowledge of this topic." (S7) Students also admired the quality submissions of their peers, and perceived reading these submissions as a learning opportunity. One student stated, "I appreciate that some of them have done a really great job. The posts were long and full of very meaningful content. I could learn more stuff by reading their posts." (S4)

The rise in eagerness to submit quality posts could be attributed to the adjustment of gamification strategies, such as the more frequently assigned quality-based badges and the display of badge rules directly on the main announcement page. These adjustments helped to emphasize the expected performance goals in this course.

The supporting pedagogical strategies (i.e., instructor elaborative feedback) empowered students to make quality submissions. One stated that, "the instructor gave us a lot of constructive feedback throughout the course, so it really helped me." (S6) She also commented that the learning design not only encouraged them to think about the completion of the activity that time, but also how they could do better next time.

Furthermore, the students made a number of suggestions, including rewarding students with real gifts or coupons (44%, n = 4) and reducing the workload by combining activities or removing tasks during the midterm exams (44%, n = 4). Two thirds of the students reported that their favorite game mechanics were the quality-based badges (66%, n = 6), such as the coffee coupons and tour package (33%, n = 3), because they were challenging to win, and the team bonus (33%, n = 3), which brought their teams closer together and made them feel rewarded. Similar to treatment group_1, six of the nine students mentioned that the sense of closeness between their teammates had been strengthened. Most importantly, 89% (eight out of nine) of the students thought that the overall learning design, i.e., the integration of gamification strategies and learning activities, was quite successful. This figure was much higher than that for treatment group_1 (50%). Overall, the students felt that the design of the learning tasks and the use of gamification strategies worked well together and ultimately facilitated their learning.

7. Discussion

The aim of this study is to test and refine a theory-based gamification design model, and examine its effects on students' learning achievement, participation, and perceptions. Specifically, we refined the GAFCC design model and applied it to three cohorts of students in an undergraduate course. The results indicated that the updated-GAFCC model were effective in enhancing learning achievements and participation, as well as motivating more students to produce uppermost level thinking artifacts. The findings contribute to our understanding of how the updated theory-based design model can be used to optimize learning outcomes.

7.1. Impact on learning achievements

In the updated-GAFCC condition, the quality of the thinking artifacts was higher than that in the GAFCC condition and the control condition. In the updated-GAFCC condition, many students used empirical data or theories to support their arguments within a clear logical framework. At the same time, creative arguments that could inspire others were frequently observed in the updated-GAFCC condition. In the GAFCC condition, more high-quality artifacts were produced than in the control group, but some students submitted low-quality artifacts (ranging from 4%-25%). This situation was much improved in the updated-GAFCC condition, with the proportion of lower quality artifacts reducing to 0% to 12%. The positive change confirmed the effectiveness of the updated-GAFCC model in promoting higher quality submissions. Research has suggested that more detailed feedback can guide the modification of students' thinking and help them understand what constitutes a quality submission (NRC, 2000). Thus, the students in the updated condition had a clearer understanding of the quality goals they needed to achieve. This result is consistent with the suggestion that gamified feedback and teacher elaborative feedback can be used together to optimize the learning outcomes in a gamified environment (Lo & Hew, 2018). At the same time, our adjustment of the gamification design, in which we presented the game rules in a more obvious section of the course page, may have raised the students' awareness of the gamified activities and promoted the internalization of certain behavioral and cognitive goals. The more timely quality-based badges also provided more immediate recognition and confirmation of students' positive behavior. Together, the adjustment of the gamification strategies and the inclusion of a supporting pedagogy motivated the students to make more in-depth submissions. In addition, more creative submissions were observed in the updated-GAFCC condition, which is consistent with the finding of the interviews that the students practiced thinking beyond the questions given by the lecturer.

The updated-GAFCC condition and the GAFCC condition performed equally well in the post-tests and post-quizzes. An interesting observation was that when completing the less challenging tasks in the post-class quizzes, such as recalling factual knowledge, the students' submissions in the two gamified conditions were of similar quality. However, when completing more complex tasks, such as in the thinking activities, the gap between the two conditions increased, with the updated-GAFCC condition stimulating more higher level critical thinking than the GAFCC condition.

7.2. Impact on task completion

More students in the updated-GAFCC condition and the GAFCC condition completed the pre-thinking activities before the due date than in the control group. The average completion rates before the due date were 29% for the control group, 62% for the gamified group, and 56% for the updated-GAFCC condition. Moreover, significantly more students in the updated-GAFCC condition and the GAFCC condition completed the post-class quizzes than in the non-gamified group in five of the seven weeks. Students in the updated-GAFCC condition explained that they felt that the content of the gamified activities was designed to enhance their knowledge and skills, and that the use of gamification strategies, such as badges and leaderboards, provided a source of positive encouragement that motivated them to participate in the activities. Moreover, the students felt they had a choice regarding whether to participate in the activities, and they did not feel a sense of being manipulated. This provided the students with a sense of control over their own learning (Deci et al., 1981) and facilitated motivational learning experiences.

7.3. Impact on perception

The interview results indicated that in the updated-GAFCC condition, 89% of the students preferred to study in a gamified environment, and that 11% would opt to learn in a gamified environment if a more competitive leaderboard was introduced. In the updated-gamified condition, the majority of the students preferred to study in a gamified environment because they could collaborate and interact more with their peers and found the experience enjoyable (Zatarain-Cabada et al., 2020). The students thought their efforts were recognized in the gamified condition (Mullins & Sabherwal, 2020), and it enhanced their sense of achievement. The high-quality thinking submissions also provided them with opportunities to learn from their peers. In sum, the students were satisfied with the overall learning design.

8. Conclusion

8.1. Impact on learning achievements

Our three-year design-based study was carried out in collaboration with the course instructor. In addition, feedback was collected from students to inform the design. This study demonstrated how the updated-GAFCC design model can be implemented in courses to engage students on the cognitive and behavioral levels. Several lessons can be learned from the design process implemented in this study: (1) it is important to build alignment between the instructional objectives, game rules, and game elements; (2) the students' sense of feeling instrumental is key to the success of a gamification design; (3) the use of gamification can enhance student learning and task completion; (4) the game rules should be displayed in a prominent place in the classroom or on the website for students to view; (5) timely elaborative feedback helps build competence; (6) lastly, iterative design and refinement are necessary for developing mature learning designs.

8.2. Limitations

This study has a number of limitations that should be noted when interpreting the results. First, we tested the updated-GAFCC design model with students enrolled in an information management course. The nature of this course may differ from that of other disciplines, such as computer programming and business education. Thus, further research is warranted on the impact of the updated-GAFCC design model in other disciplines. Second, the sample used in this study was only representative of Asian students in a higher education setting, and the learner characteristics and user types may have influenced their acceptance of the gamification strategies and the overall learning design. Moreover, it should be noted that the interviews were conducted on a voluntary basis. Therefore, the students who participated in the interviews were more likely to have been those who had higher motivation levels (Robinson, 2014), and so we have to be cautious with regards to the generalizability of the collected information. In the future, it would be exciting to apply the updated design model in other cultures and compare the results across settings.

Acknowledgement

The research was supported by a grant from the Research Grants Council of Hong Kong (Project reference no: 17616020).

References

Aguilar, S. J., Holman, C., & Fishman, B. J. (2018). Game-inspired design: Empirical evidence in support of gameful learning environments. *Games and Culture*, *13*(1). 44-70.

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational Technology and Society*, *11*(4), 29-40.

Bai, S., Hew, K. F., & Huang, B. (2020). Does gamification improve student learning outcome? Evidence from a meta-analysis and synthesis of qualitative data in educational contexts. *Educational Research Review*, *30*, doi:10.1016/j.edurev.2020.100322

Barata, G., Gama, S., Jorge, J., & Goncalves, D. (2017). Studying student differentiation in gamified education: A Long-term study. *Computers in Human Behavior, 71*, 550-585.

Baydas, O., & Cicek, M. (2019). The Examination of the gamification process in undergraduate education: A Scale development study. *Technology, Pedagogy and Education*, *28*(3), 269-285.

Bolboacă, S. D., Jäntschi, L., Sestras, A. F., Sestras, R. E., & Pamfil, D. C. (2011). Pearson-Fisher Chi-Square statistic revisited. *Information*, *2*(3), 528-545.

Cai, F., Dai, G., & Han, T. (2016). Gamification design based research on fitness mobile application for university students. In A. Marcus (Ed.), *Design, user experience, and usability: Novel user experiences*. *(pp.240-251).* Cham, Switzerland: Springer.

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews: Problems of Unitization and Intercoder Reliability and Agreement. *Sociological Methods & Research*, *42*(3), 294–320.

Csikszentmihalyi, M. (1990). *Flow: The Psychology of optimal experience*. New York, NY: Harper & Row.

Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry, 11*(4), 227-268.

Deci, E. L., Nezlek, J., & Sheinman, L. (1981). Characteristics of the rewarder and intrinsic motivation of the rewardee. *Journal of Personality and Social Psychology*, *40*(1), 1-10.

Deterding, S., Khaled, R., Nacke, L. E. & Dixon, D., (2011). From Game Design Elements to Gamefulness: Defining "Gamification". *In Proceedings of the 15th International Academic MindTrek Conference on Envisioning Future Media Environments* (pp.9-15). New York, NY:: Association for Computing Machinery.

Dichev, C., & Dicheva, D. (2017). Gamifying education: What is known, what is believed and what remains uncertain: A Critical review. *International Journal of Educational Technology in Higher Education, 14*(9), doi:10.1186/s41239-017-0042-5

Ding, L. (2019). Applying gamifications to asynchronous online discussions: A Mixed methods study. *Computers in Human Behavior*, *91*, 1-11.

Festinger, L. (1954). A Theory of social comparison processes. *Human Relations, 7*(2), 117-140.

Glaser, B. G. (1965). The Constant comparative method of qualitative analysis. *Social Problems*, *12*(4), 436–445.

Greenlaw, S. A., & DeLoach, S. B. (2003). Teaching critical thinking with electronic discussion. *The Journal of Economic Education, 34*(1), 36-52.

Huang, B., & Hew, K. F. (2018). Implementing a theory-driven gamification model in higher education flipped courses: Effects on out-of-class activity completion and quality of artifacts. *Computers & Education*, *125*, 254-272.

Huang, B., Hew, K. F., & Lo, C. K. (2019). Investigating the effects of gamification-enhanced flipped learning on undergraduate students' behavioral and cognitive engagement. *Interactive Learning Environments*, *27*(8), 1106-1126.

Jo, J., Jun, H., & Lim, H. (2018). A Comparative study on gamification of the flipped classroom in engineering education to enhance the effects of learning. *Computer Applications in Engineering Education*, *26*(5), 1626-1640. doi:10.1002/cae.21992

Jong, M. S. Y. (2019). Sustaining the adoption of gamified outdoor social enquiry learning in high schools through addressing teachers' emerging concerns: A 3-year study. *British Journal of Educational Technology*, *50*(3), 1275-1293.

Krippendorff, K. (2004). *Content analysis: An Introduction to its methodology* (2nd ed.). Thousand Oaks, CA: Sage.

Kyewski, E., & Kramer, N. C. (2018). To gamify or not to gamify? An Experimental field study of the influence of badges on motivation, activity, and performance in an online learning course. *Computers & Education, 118*, 25-37.

Landers, R. N., Bauer, K. N., & Callan, R. C. (2017). Gamification of task performance with leaderboards: A Goal setting experiment. *Computers in Human Behavior*, *71*, 508-515.

Lee, J. J., Matamoros, E., Kern, R., Marks, J., de Luna, C., & Jordan-Cooley, W. (2013). Greenify: Fostering sustainable communities via gamification. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 1497-1502). New York, NY: Association for Computing Machinery.

Li, W., Grossman, T., & Fitzmaurice, G. (2012). GamiCAD: A Gamified tutorial system for first time Autocad users. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology* (pp. 103-112). New York, NY: Association for Computing Machinery.

Lo, C. K., & Hew, K. F. (2018). A Comparison of flipped learning with gamification, traditional learning, and online independent study: The effects on students' mathematics achievement and cognitive engagement. *Interactive Learning Environments*, *28*(4), 464-481.

Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, *57*(9), 705-717.

Lopez, C. E., & Tucker, C. S. (2019). The Effects of player type on performance: A Gamification case study. *Computers in Human Behavior*, *91*, 333-345.

McKenney, S., & Reeves, T. (2012). *Conducting educational design research*. New York, NY: Routledge.

Mullins, J. K., & Sabherwal, R. (2020). Gamification: A Cognitive-emotional view. *Journal of Business Research, 106*, 304-314.

National Research Council. (2000). *How people learn: Brain, mind, experience, and school: Expanded edition*. Washington, DC: The National Academies Press. doi:10.17226/9853

Rigby, C. S., & Ryan, R. M. (2011). *Glued to games: How video games draw us in and hold us spellbound*. Santa Barbara, CA: Praeger.

Robinson, O. C. (2014). Sampling in interview-based qualitative research: A Theoretical and practical guide. *Qualitative Research in Psychology*, *11*(1), 25-41.

Sailer, M., & Sailer, M. (2020). Gamification of in-class activities in flipped classroom lectures. *British Journal of Educational Technology*. doi:10.1111/bjet.12948

Skinner, B. F. (1953). *Science and human behavior*. New York, NY: Macmillan.

Yildirim, I. (2017). The Effects of gamification-based teaching practices on student achievement and students' attitudes toward lessons. *The Internet and Higher Education*, *33*, 86-92.

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

Zatarain-Cabada, R., Barrón Estrada, M. L., Ríos Félix, J. M., & Alor Hernández, G. (2020). A Virtual environment for learning computer coding using gamification and emotion recognition. *Interactive Learning Environments*, *28*(8), 1048-1063.