

Gamification of Assessment Test through Multiple Question Paths to Facilitate Participants' Autonomy and Competence

Pratama Wirya Atmaja ^{a,1*}, Eka Prakarsa Mandyartha ^{a,2}

^a Department of Informatics, University of Pembangunan Nasional "Veteran" Jawa Timur
Jalan Raya Rungkut Madya, Gunung Anyar, Surabaya

¹ pratama_wirya.fik@upnjatim.ac.id; ² eka_prakarsa.fik@upnjatim.ac.id

Article Info

Article history:

Received: March 14, 2020

Revised: April 7, 2020

Accepted: April 18, 2020

Keyword:

Gamification

Multiple question path

Assessment test

Autonomy

Competence

ABSTRACT

Gamifying activities to make them more game-like is one of the hottest trends in various fields, including education. Among the factors influencing the success of gamification for education are the participants' sense of autonomy and competence, which can be facilitated with the incorporation of multiple learning paths. However, the use of multiple question paths in gamified assessment tests is still under-studied. This mixed-method study was aimed at exploring the matter through a paper-based and gamified assessment test in higher education. A controlled experiment was conducted in a calculus course in an informatics department. The experimental group (n = 38) undertook a gamified written test, and the control group (n = 37) undertook a regular one. The gamified test consisted of several Hard and Medium Questions, and each participant would choose a question path containing some of the questions. Nine question paths were available with varying ratios between Hard and Medium Questions, and the participants were allowed to ask for two hints on the Hard ones. A questionnaire, based on the EGameFlow model, was used to assess the gamified test. The results show that the gamified test was able to facilitate the participants' sense of autonomy but not their sense of competence, which was due to flaws of the test. Two additional positive effects of the test on the participants' knowledge improvement and Flow experience are identified. The path selection pattern among the participants and the flaws of the gamified test are also discussed.

I. INTRODUCTION

Gamification is a rapidly-expanding method and technology, which can be defined as the use of game elements in activities that, in essence, are unrelated to games [1]. The main goal of gamification is to increase the fun factor of the activities and make them less dull in the eyes of the participants. The applications of gamification have spread widely into numerous fields in the society, such as primary education [2], higher education [3], tourism [3][4], marketing [5][6], recruitment [6][7], and even research [8] and software engineering [9].

Due to its cutting-edge nature, it is no wonder that the applications of gamification have been mostly found in developed countries. On the other hand, those with less powerful economies are slowly but surely embracing it. Halloluwa et al. [10] studied the use of tablet-based gamification for primary school students in Sri Lanka. Ezezika et al. [11] conducted a pilot study on gamification of nutrition for Nigerian adolescents. Udjaja et al. [12] developed an application for gamified elementary math teaching in Indonesia, which was enhanced with multimedia elements. Utomo and Santoso [13] developed a gamification-based pedagogical agent with human-like expressions, which could

provide personalized feedbacks to e-learning users in Universitas Indonesia.

Gamification is commonly applied digitally, although non-digital applications are also possible, and even feasible, in some contexts [14]. Taspinar et al. [15] applied gamification through a board game to support blended learning in higher education. Indeed, non-digital games are still a common sight in the field of education, especially as some people find digital games to be uncomfortable to use [16].

Over the years, many aspects of education have been gamified, which includes assessments. In addition to being more enjoyable than their non-gamified counterparts, Kocadere and Çağlar [17] found that gamified assessment processes also stimulate motivations, induce Flow, and decrease tension and anxiety among the participants. Moccozet et al. [18] inserted gamification elements into the process of group task assessment to encourage students to contribute to their groups. Zainuddin et al. [19] applied gamified e-quizzes in science classes, which were effective in assessing students' comprehension. Attali and Arieli-Attali [20] observed that game points encourage assessment participants to answer questions more quickly. Guerreiro and Nordengren [21] noted that game-like features in assessment increase student engagement, which in turn drives the students to give their best efforts.

The promise of gamification (turning just about anything from our everyday lives into a fun and engaging activity) sounds fascinating; however, many of its applications have been proven ineffective [22][23][24]. What, then, are the necessary ingredients of successful gamification? According to the state of the art [25][26][27], among them are the three elements of the Self-Determination Theory (SDT): relatedness, autonomy, and competence. Relatedness is about the social aspect of learning, which makes a participant feel that they belong in the group of like-minded learners. Autonomy is about the participant's free will and freedom in learning things that suit their needs. Competence is about the participant's success in the learning process, especially compared to what their peers are able to achieve.

Relevant to this research are the second and the third element. According to Nicholson [28] and Aparicio et al. [29], autonomy can be facilitated through several gamification features, one of them being multiple learning paths that each participant can choose at will. The feature is also in line with another established theory, the Universal Design for Learning [30], which emphasizes multiple means for students in learning and mastering an educational topic. On the other hand, features such as balanced levels of challenge (neither too hard nor too easy) can facilitate the sense of competence.

Many authors have explored the matter of facilitating autonomy and competence in educational gamification. Böckle et al. proposed a design framework for adaptive gamification, with one of its many design principles being "design multiple paths (choices) to achieve end-user goals and support their believe and motivation" [31]. Gordon et al. [32] conducted a longitudinal study on mathematics assessment with several game-like features, one of them being adaptive difficulty. Jagušt et al. [33] utilized a personalized adaptive

algorithm to match the presented questions to each student's skill level. Iosup and Epema [34] applied gamification in technical undergraduate and graduate courses, which provided multiple paths for students of the "explorer" type. Aguilar et al. [35] conducted gamified courses at a research university, which supported the participants' autonomy by letting them choose and customize their assignments.

Multiple paths and balanced challenges are also recognized as two of the fundamental elements of game player experience [36]. Both are closely related: one can choose specific paths with challenges appropriate to one's skill level. With regards to assessment, a research gap can be identified: the application of multiple paths in gamified assessment tests is still under-explored. The only work on the topic we are aware of was by Guerreiro and Nordengren [21], who implemented gamified assessment prototypes with adventure game-like features, including multiple paths for the participants to choose. However, the authors did not explain the technical details of the paths, including whether the path choices affected the assessment itself or only the adventure game-like experience.

This study was then aimed to fill the gap and shed light on the topic. It was done by means of a controlled experiment on gamified written tests with multiple question paths. Data collection and analysis were done in a mixed-method manner.

II. RESEARCH QUESTIONS

Based on the research gap, several research questions were formulated as follows:

RQ1: Can multiple question paths facilitate a sense of autonomy among the test participants?

RQ2: Can multiple question paths facilitate a sense of competence among the test participants?

RQ3: Are there other positive effects of the multiple question paths to the participants?

RQ4: What are the pattern of path selection among the participants?

RQ5: What are the difficulties in implementing multiple question paths in an assessment test?

III. RESEARCH METHODOLOGY

A controlled experiment was conducted to answer the five research questions. The experiment applied a gamified assessment test to the experimental group and a regular one to the control group. The experiment was conducted in a real higher education setting in the Department of Informatics, Faculty of Computer Science, University of Pembangunan Nasional (UPN) "Veteran" Jawa Timur.

The gamification of the experimental group's test was done in an analog manner. Both the gamified and non-gamified tests were paper-based, with the gamified one requiring more paper sheets for the gamification resources.

A. Specifications of the Assessment Tests

Both the gamified and non-gamified assessment tests were on calculus for informatics students. The tests assessed the participants' comprehension of derivatives and integrals. The non-gamified test consisted of seven questions, whereas the gamified one consisted of 20 questions. The goal of every participant of both tests was to try to get 100 points. As the total points of the non-gamified test's seven questions were 100, its participants had to answer all of the questions, or else they would get less-than-perfect scores. The participants of the gamified test, on the other hand, had much more freedom in answering the questions, as explained in the next subsection.

Two calculus classes participated in the experiment; the first class became the experimental group, and the second class became the control one. The calculus tests were conducted in classrooms in the department, as parts of the final examination week of the semester. As both classes were scheduled at separate days in the examination week, their tests were required to consist of different questions. The experimental group undertook the gamified test first, followed by the control group several days later.

B. Design of the Gamified Assessment Test

The gamified test consisted of nine derivative questions and 11 integral questions. Figure 1 shows the question path map of the gamified test, which was given to every participant on a sheet of paper. Difficulty-wise, the 20 questions were split into Hard Questions and Medium Questions. Each Medium Question was worth 10 points, whereas each Hard Question was worth 20 points. Because the Hard Questions were significantly harder than the Medium ones, the participants were allowed to ask for hints on the Hard Questions. The hints were written on small sheets of paper and would guide the participants without revealing too much about the correct answers. A participant could ask for hints on at most two Hard Questions. Figure 2 shows an example of a Hard Question and its hint.

In summary, there were two game-like mechanics in the gamified test: the multiple question paths, being the primary mechanic, and the limited-hint system, being the secondary one. Both were applied for their simplicity and fun factors. Both were not hard to prepare (to the assessors) and understand (to the participants). As noted by Miller et al. (2014), making the participants understand, as quickly as possible, how to interact with the gamified system is essential in minimizing their frustrations. As the experimental group participants were briefed on the gamified test just before they undertook it, the simplicity of the mechanics played a vital role in ensuring the test ran smoothly.

Each participant was free to choose a path that best suited their taste and skill. They could pick a path with many Medium Questions and only a few Hard Questions and vice versa. There was one restriction: they were not allowed to switch to another path once they had tried to answer at least one question in the selected path.

In total, there were nine question paths in the gamified test, and each rewarded a total of 100 points to the participants. From the first to the last path, each consisted of:

- 1) Node 1, 2, 5, 11, 14, and 19;
- 2) Node 1, 2, 5, 8, 12, 14, and 19;
- 3) Node 1, 2, 5, 8, 12, 15, 17, and 19;
- 4) Node 1, 2, 6, 9, 13, 17, and 19;
- 5) Node 1, 2, 6, 9, 13, 16, 18, and 20;
- 6) Node 1, 2, 6, 7, 10, 13, 17, and 19;
- 7) Node 1, 2, 6, 7, 10, 13, 16, 18, and 20;
- 8) Node 1, 2, 3, 4, 7, 10, 13, 17, and 19;
- 9) Node 1, 2, 3, 4, 7, 10, 13, 16, 18, and 20.

Among the paths, path one was the shortest, and path nine was the longest. On the other hand, path 1 had the most number of Hard Questions (which was four), whereas path nine was composed of only Medium Questions.

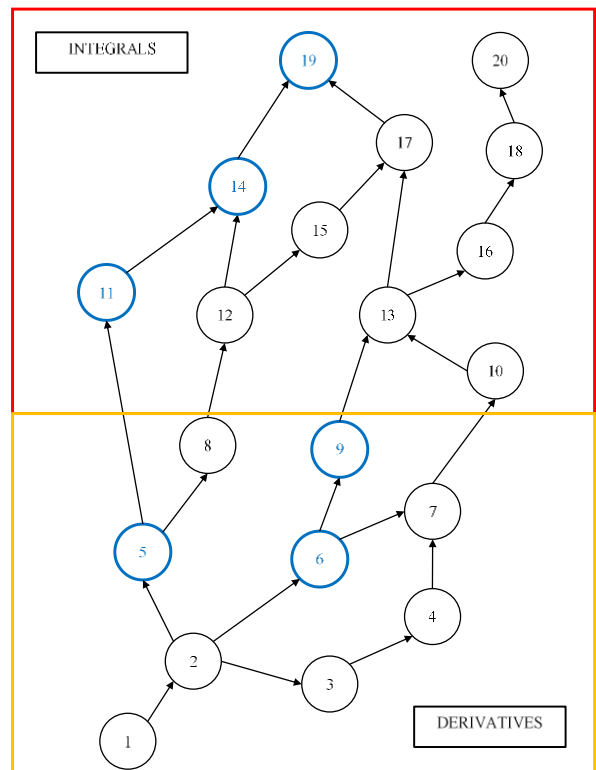
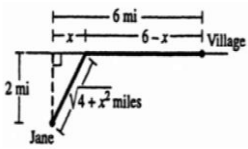


Fig. 1. The map of the gamified calculus test. The questions in the bottom box are about derivatives, and the ones in the top box are about integrals. Every participant started at the first node and went on to reach one of the last nodes (the 19th or the 20th node). Blue nodes are the Hard Questions, whereas the black ones are the Medium Questions.

Dul Kribo is 2 mi offshore in a boat. He wishes to reach a coastal village 6 mi down a straight shoreline from the point nearest the boat. He can row 2 mph and can walk 5 mph. Where should he land his boat to reach the



Dul Kribo's travel distance will be $\sqrt{4 + x^2} + (6 - x)$. What is asked here is his travel time, which equals the distance divided by his travel speed. Keep in mind that his rowing and walking speed are different

Fig. 2. An example of a Hard Question (top box) and its hint (bottom box). The Hard Questions were adapted from the 13th edition of *Thomas' Calculus: Early Transcendentals* book.

C. Ensuring Fairness between Experimental and Control Groups

The non-gamified test was prepared as a regular written test on calculus. Since it did not utilize a hint system, it would be unfair if it included questions as advanced as the Hard Questions in the gamified one. Therefore, the question set in the non-gamified test was made more or less equal to the Hard Question-less path (path 9) in the gamified test to ensure fairness between the groups.

D. Experiment Participants

Seventy-five first-year students of the Informatics Department of UPN "Veteran" Jawa Timur participated in the experiment. Thirty-eight students formed the experimental group, and 37 others formed the control group.

E. Data Collection and Means of Answering the Research Questions

Both quantitative and qualitative data were collected and analyzed. Two types of quantitative data were collected during the experiment and analyzed: (1) assessment test questionnaire results; and (2) the experimental group participants' path selections. Table 1 shows the assessment questionnaire used to gather the first data. The data were analyzed through three means: (1) mean values and standard deviations; (2) differences between the two groups' results; and (3) correlation analysis. The first and second means will be used to answer RQ1 and RQ2, whereas RQ3 will be answered through the third mean.

The second data were collected by analyzing the experimental group participants' answer sheets and will be used to answer RQ4. The qualitative data from a post-test discussion with the experimental group participants will be used to answer RQ5. Meanwhile, the test scores of the participants were not analyzed. The reason for the exclusion was that the gamified and non-gamified tests presented entirely different questions, so the scores were not directly comparable between the groups.

F. Assessment Test Questionnaire

The assessment test questionnaire was based on the EgameFlow model [38], which is widely used to assess the quality of games for learning. The model is composed of eight dimensions, each consisting of several assessment items. Table 1 shows the items of the questionnaire, alongside the five dimensions and eight items of EgameFlow that became the bases for the questionnaire items. Questionnaire item no. 9 was an exception as it was used to assess the experimental group participants' test-type preferences and, therefore, was not based on EGameFlow.

Of the eight EGameFlow-based items, items no. 7 and 5 are directly related to autonomy and competence, respectively, whereas item no. 6 is related to the multiple-path mechanic. The three items will be used to answer RQ1 and RQ2. Meanwhile, items no. 1, 2, 3, 4, 8, and 9 were used to assess various other positive experiences of the participants during the tests. An analysis of possible correlations between the first three items (no. 5, 6, and 7) and the other six will be used to answer RQ3 and corroborate the answer to RQ1 and RQ2.

The participants of both the experimental and control groups filled the questionnaire, with two exceptions: item no. 6 and 9 were not presented to the control group, as the items were not related to the non-gamified test.

IV. RESULTS AND DISCUSSIONS

Table 2 shows the results of the questionnaire. Through a Kolmogorov-Smirnov Test, it was found that none of the results were normally distributed, with the exception of that of item no. 9 (which was not presented to the control group). Therefore, a two-tailed Mann-Whitney U test was then performed on the results to get the differences between the groups. For the correlation analysis, Spearman Correlation was also used.

A. Questionnaire Results

The questionnaire results are reliable, with Cronbach's Alpha value of 0,7925 for those of the gamified test and 0,7568 for those of the non-gamified one. It can be seen that the questionnaire results of the gamified test are better than those of the non-gamified one in four items (in light green), whereas the opposite happens in three items (in light brown). The gamified test acquires a better score than its counterpart in item no. 7, which is directly related to autonomy. On the other hand, the opposite happens to the gamified test's item no. 5 score, which is slightly lower than that of the non-gamified one. It seems that the gamified test was successful in facilitating autonomy, yet fell short in facilitating competence. Looking at the Mann-Whitney U test results, however, the differences between the groups' results are insignificant. Regardless of the insignificance, the correlation analysis in the next subsection corroborated the findings related to the sense of autonomy and competence.

TABLE I. ASSESSMENT TEST QUESTIONNAIRE

EGameFlow Dimension	EGameFlow Item	Item no.	Questionnaire Item
Knowledge Improvement	The game increases my knowledge.	1	You feel that the calculus test helped you understand calculus.
	I want to know more about the knowledge taught.	2	You feel more enthusiastic about learning calculus after undertaking the calculus test.
Concentration	Generally speaking, I can remain concentrated in the game.	3	During the calculus test, you could concentrate on answering the questions.
	I am not burdened with tasks that seem unrelated.	4	During the calculus test, you did not feel distracted by unrelated tasks.
Challenge	Workload in the game is adequate.	5	The difficulty and workload of the test were appropriate for you.
	The game provides different levels of challenge that tailor to different players.	6	During the calculus test, you were able to adjust the difficulty levels of the questions to suit your needs.
Autonomy	I feel a sense of control over the game.	7	During the calculus test, you felt that you could freely control how you undertook it.
Immersion	I forget about time passing while playing the game.	8	During the calculus test, you felt engaged and lost track of time.
		9	You prefer the gamified test over the regular calculus test.

Therefore, RQ1 and RQ2 can be answered as follows:

- The multiple-path mechanic can, indeed, facilitate the participants' sense of autonomy;
- The multiple-path mechanic can also *possibly* facilitate the participants' sense of competence if the mechanic successfully helps them selecting a path with the right difficulty level (this is explained further in the next subsections).

B. Correlation Analysis Results

Tables 3 and 4 show the results of correlation analysis conducted on the questionnaire results of the experimental and the control group, respectively. Akoglu discussed three conventions for interpreting the strength of a bivariate correlation coefficient [39]. The Dancey & Reidy convention, which came from the field of psychology, is used here to interpret the Spearman's rho values in Tables 3 and 4.

Within the experimental group, the multiple-path mechanic, assessed with item no. 6, seems to correlate moderately with the participants' sense of autonomy, assessed with item no. 7, but not with their sense of competence, assessed with item no. 5. There also seems to be a weak correlation between the participants' sense of competence and their sense of autonomy. This interesting finding indicates that the application of the mechanic might be flawed but not entirely so.

Why was not the multiple-path mechanic successful in facilitating a sense of competence? One explanation for that is that selecting a question path was, ultimately, not beneficial to the participants. Each of them wanted to undertake a path with the right difficulty level for them, but identifying and choosing such a path might be harder than it seemed. This problem will be discussed further in the "Flaws of the Gamified Test" subsection.

The disconnect between the sense of autonomy and competence among the participants does not seem to have happened in the control group. The results of their fifth and seventh items are correlated moderately, which suggests that,

without the hindrance of game-like mechanics that do not benefit them, the pairing of autonomy and competence may come naturally to a participant.

The correlations between items no. 5 to 7 and item no. 9, which assessed the participants' preferences toward the gamified test, further corroborate the failure of the gamified test in facilitating competence. The correlations with items no. 6 and 7 are weak, whereas that with item no. 5 is moderate. It may mean that the participants were generally indifferent toward the gamified test and its promise of facilitating autonomy. The indifference then led them to adopt a rather pragmatic attitude: "If it helps me answering the questions and getting a good score, then it is good; otherwise, I do not care about it."

Meanwhile, the participants' post-test enthusiasm toward the assessed subject, as represented by item no. 2, correlates only weakly with items no. 5 to 7 in the experimental group. On the other hand, the knowledge gained from the test, represented by item no. 1, correlates moderately with the result of item no. 6 in the group. It seems that, during selecting and solving question paths, the gamified test participants were able to learn something new on the assessed subject. It is quite likely that the limited-hint system played a major part in the learning effect, as the hints guided the participants in composing their Hard Question answers. It is curious, however, that a post-test effect on the participants' learning enthusiasm did not follow the in-test effect on their knowledge. The reason for that is not yet apparent.

In the control group, another peculiar phenomenon can be observed: the result of item no. 2 correlates moderately with that of item no. 7 but only weakly with that of item no. 5. It may mean that a participant will want to continue their pursuit of knowledge if they feel in control of the pursuit (i.e., having learning autonomy), regardless of their performance in it (i.e., their competence). However, it is not clear why such a thing did not also happen in the experimental group.

Of all the positive experiences of the gamified test participants, the feeling of losing track of time (item no. 8) is the one with the strongest and most decisive positive correlations with their sense of autonomy, competence, and the multiple-path mechanic. The feeling, together with the

sense of autonomy and competence, are three elements of the highly sought-after Flow experience [39]. The moderately-strong correlations between them in the experimental group hint at the Flow experience among the group participants and that the multiple-path mechanic played a part in it. In the control group, a correlation of similar strength also exists between the results of items no. 8 and 5 (competence) but, curiously, not between those of items no. 8 and 7 (autonomy). The reason can only be speculated: the lack of support for the participants' autonomy in the non-gamified test drove the participants not to associate their sense of autonomy with their Flow experience.

TABLE II. QUESTIONNAIRE RESULTS OF EXPERIMENTAL AND CONTROL GROUPS

Item no.	Experimental Group		Control Group		Mann-Whitney U Test	
	Mean	St. Dev.	Mean	St. Dev.	p-value	Significant
1	3.55	0.86	3.59	0.72	0.865	No
2	3.76	0.75	3.43	0.83	0.129	No
3	3.74	0.72	3.43	0.77	0.056	No
4	3.84	0.82	3.86	0.71	0.992	No
5	3.21	0.78	3.24	1.01	0.638	No
6	3.84	0.89				
7	3.95	0.80	3.62	0.92	0.147	No
8	3.79	1.14	3.65	1.03	0.447	No
9	3.89	0.95				

TABLE III. SPEARMAN CORRELATIONS BETWEEN QUESTIONNAIRE RESULTS OF EXPERIMENTAL GROUP

Item Pair		Spearman's Rho	p-value	Significant
1st Item	2nd Item			
5	1	0.29021	0.07715	No
	2	0.36804	0.02300	Yes
	3	0.08282	0.62108	No
	4	0.15569	0.35061	No
	6	0.26168	0.11250	No
	7	0.17672	0.28853	No
	8	0.50732	0.00115	Yes
	9	0.40524	0.01161	Yes
6	1	0.42405	0.00797	Yes
	2	0.35690	0.02784	Yes
	3	0.27771	0.09139	No
	4	-0.10354	0.53616	No
	7	0.54002	0.00047	Yes
	8	0.55924	0.00026	Yes
	9	0.16698	0.31634	No
	1	0.33240	0.04145	Yes
7	2	0.33686	0.03864	Yes
	3	0.09580	0.56721	No
	4	0.08838	0.59773	No
	8	0.52180	0.00078	Yes
	9	0.12054	0.47100	No

TABLE IV. SPEARMAN CORRELATIONS BETWEEN QUESTIONNAIRE RESULTS OF CONTROL GROUP

Item Pair		Spearman's Rho	p-value	Significant
1st Item	2nd Item			
5	1	0.36362	0.02695	Yes
	2	0.28038	0.09280	No
	3	0.16854	0.31868	No
	4	0.20751	0.21780	No
	7	0.49017	0.00207	Yes
	8	0.52052	0.00096	Yes
7	1	0.35821	0.02948	Yes
	2	0.43241	0.00752	Yes
	3	0.29499	0.07632	No
	4	0.22748	0.17573	No
	8	0.39346	0.01599	Yes

TABLE V. NUMBER OF PARTICIPANTS SELECTING EACH QUESTION PATH

Path no.	Number of Hard Questions	Number of Participants
1	4	2
2	3	0
3	2	2
4	3	12
5	2	14
6	2	3
7	1	0
8	1	1
9	0	4
Total Number of Participants		38

The last two items are no. 3 (able to concentrate) and 4 (not being distracted), the results of which correlate *decisively weakly* with the results of items no. 5 to 7 in both groups. In the experimental group, the results of item no. 4 and 6 are even *negatively* correlated, albeit insignificantly so statistic-wise. Given the multiple-path mechanic's failure in facilitating competence, it should come as no surprise that some participants thought of it as a mere distraction. In summary, the correlation analysis has answered RQ3 as follows:

- The multiple-path mechanic, together with the limited-hint system, can improve each gamified test participant's knowledge of the assessed subject;
- The multiple-path mechanic can (indirectly, at least) make the participants lose track of time, which is a prerequisite of the Flow experience.

C. Distribution of Path Selections among Participants

Table 5 shows the distribution of path selections among experimental group participants. The most popular paths were those with two or three Hard Questions, which were path 4 and path 5. On the other hand, the one with no Hard Questions, which was path 9, was selected by only four participants. One possible cause of the pattern of path selection is the participants' curiosity about the gamified test, which drove them to engage with the Hard Questions and the two-hint-at-most system. On the other hand, they also wanted to take advantage of the system to maximize their scores; therefore, most of them avoid the path with the most number of Hard Questions (path 1).

Meanwhile, path 1 was selected by only two participants, who were among the most successful academic-wise in their class. It should not be surprising, as taking on path 1 required the participants to be quite confident in their skills.

In summary, the analysis of path selection pattern can answer RQ4 as follows: Which paths will be most frequently selected depends on two non-mutually exclusive factors: (1) the benefits of the paths; and (2) whether the paths satisfy the participants' curiosity about the game-like mechanics.

D. Flaws of the Gamified Test

The post-test discussion with the participants uncovered two flaws of the gamified test. The flaws reflect the practical difficulties in conducting a gamified test with a multiple-path mechanic.

The first flaw was in the classification of questions according to their actual difficulty levels. Some participants felt that some of the Medium Questions were harder than the others. It made them regret selecting paths with Medium Questions that were harder than those in other paths. The failure in assessing the questions' difficulty levels was hard to avoid, as the participants often realized the actual difficulties after they had spent some time to deal with the questions. As a side effect, they also complained about the rule prohibiting them from switching paths.

The second flaw was about the appropriateness of the hints for the Hard Questions. Some participants felt that the hints they received were not useful in guiding their answers. On the other hand, some other participants admitted that their hints helped them so much, to the point that they felt they had unfair advantages over their classmates who chose other question paths. On the flip side, this flaw was not as critical as the first one, as the hint system could be replaced with another game-like mechanic, as long as the mechanic would be equally fun and easy to understand.

In summary, RQ5 can be answered as follows. The primary difficulty in conducting an assessment test with multiple question paths is in ensuring the difficulty level balance in the paths. The secondary difficulty is in supporting the multiple-path mechanic with other mechanics that increase the overall participant engagement without harmful side effects.

V. LIMITATIONS

This research is limited in two ways. First, the relatively small sample size limits the generalization of the findings. Second, the incorporation of the limited-hint mechanic in the gamified test means that the findings may not apply when a different set of game-like mechanics is utilized. It is especially relevant in a gamified digital test, as the range of possible mechanics to implement will be significantly more extensive than in paper-based ones. One possible difference will be in the answers to RQ3: without the limited-hint system or another similar mechanic, the in-test knowledge improvement effect may not happen.

A. Threats to Validity

Three potential threats to the validity of the results can be identified. First, the different sets of questions for the gamified and non-gamified test might have also influenced the questionnaire results. Innovative learning with technology development is an important part of improving the quality of learning [40]. As previously explained, efforts were made to ensure that both tests were fair, with regards to each test's characteristics. Despite the efforts, it was still entirely possible for one question set to be unfairly harder or easier than the other.

Second, the fact that the gamified and non-gamified tests were held at different days, and that the gamified one was held first, may also be considered as a validity threat. Between the end of the gamified test and the start of the non-gamified one, the experimental group participants had an ample opportunity to inform their friends in the control group about the gamified test. The information might then influence the control group participants to judge the regular test more negatively.

Third, the experiment being held in classrooms as a part of the final examination might have allowed a validity threat to occur in the form of cheatings by the participants. Efforts were made to ensure the paper-based tests were conducted per the faculty rules; however, it was still possible for the students to cheat by stealthily conversing and sharing their answers and Hard Question hints. The cheatings, then, might influence the path selection pattern (due to the cheating participants taking on the same paths as those sitting next to them so that they could solve the paths together).

VI. CONCLUSIONS

The use of multiple, selectable paths to facilitate autonomy and competence have been acknowledged in the literature of gamification for education; however, gamified assessment tests employing multiple question paths are still a strange concept. This paper has presented a mixed-method study to explore the matter, which was through a controlled experiment involving gamified and non-gamified calculus tests in higher education. The gamified test employed a multiple-path mechanic, which allowed the participants to select one from nine possible question paths with different difficulty levels, which were determined by how many Hard and Medium Questions were in each path.

A secondary game-like mechanic in the form of a limited-hint system was also employed to help the participants answer the Hard Questions. The results show that the multiple-path mechanic was able to facilitate autonomy but not competence, which was due to flaws in the gamified test. Two additional positive effects of the gamified test have been identified, and the flaws of the gamified test and the path selection pattern among the participants have also been discussed. Overall, the results imply the potentials of the multiple-path mechanic in increasing test participants' engagement.

Some fruitful topics for future researches are (1) investigating the design principles and framework for multiple question paths in assessment tests; (2) exploring the combinations of the multiple-path mechanic with other

complimenting mechanics; and (3) exploring ways to help teachers or assessors assess the difficulties of question nodes correctly and balance the question paths.

References

- [1] K. Huotari and J. Hamari, "A definition for gamification: anchoring gamification in the service marketing literature," *Electron. Mark.*, vol. 27, no. 1, pp. 21–31, 2017.
- [2] F. Garcia-Sanjuan, S. Jurdi, J. Jaen, and V. Nacher, "Evaluating a tactile and a tangible multi-tablet gamified quiz system for collaborative learning in primary education," *Comput. Educ.*, vol. 123, pp. 65–84, 2018.
- [3] S. Subhash and E. A. Cudney, "Gamified learning in higher education: A systematic review of the literature," *Comput. Human Behav.*, vol. 87, pp. 192–206, 2018.
- [4] F. Xu, F. Tian, D. Buhalis, J. Weber, and H. Zhang, "Tourists as Mobile Gamers: Gamification for Tourism Marketing," *J. Travel Tour. Mark.*, vol. 33, no. 8, pp. 1124–1142, 2016.
- [5] C. F. Hofacker, K. de Ruyter, N. H. Lurie, P. Manchanda, and J. Donaldson, "Gamification and Mobile Marketing Effectiveness," *J. Interact. Mark.*, vol. 34, pp. 25–36, 2016.
- [6] C. K. Lo and K. F. Hew, "A comparison of flipped learning with gamification, traditional learning, and online independent study: the effects on students' mathematics achievement and cognitive engagement," *Interact. Learn. Environ.*, 2018.
- [7] M. B. Armstrong, R. N. Landers, and A. B. Collmus, "Gamifying recruitment, selection, training, and performance management: Game-thinking in human resource management," in *Emerging Research and Trends in Gamification*, 2015, pp. 140–165.
- [8] J. Barwick, D. Watkins, E. Kirk, and E. Law, "Adventures with Lex: The gamification of research?," *Convergence*, vol. 24, no. 3, pp. 229–250, 2018.
- [9] T. Dal Sasso, A. Mocci, M. Lanza, and E. Mastrodicasa, "How to gamify software engineering," in *SANER 2017 - 24th IEEE International Conference on Software Analysis, Evolution, and Reengineering*, 2017.
- [10] T. Halloluwa, D. Vyas, H. Usoof, and K. P. Hewagamage, "Gamification for development: a case of collaborative learning in Sri Lankan primary schools," *Pers. Ubiquitous Comput.*, vol. 22, no. 2, pp. 391–407, 2018.
- [11] O. Ezezika, J. Oh, N. Edeagu, and W. Boyo, "Gamification of nutrition: A preliminary study on the impact of gamification on nutrition knowledge, attitude, and behaviour of adolescents in Nigeria," *Nutr. Health*, vol. 24, no. 3, pp. 137–144, 2018.
- [12] Y. Udjaja, V. S. Guizot, and N. Chandra, "Gamification for Elementary Mathematics Learning in Indonesia," *Int. J. Electr. Comput. Eng.*, vol. 8, no. 5, pp. 3860–3865, Oct. 2018.
- [13] A. Y. Utomo and H. B. Santoso, "Development of gamification-enriched pedagogical agent for e-learning system based on community of inquiry," in *ACM International Conference Proceeding Series*, 2015, pp. 1–9.
- [14] 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*, MindTrek 2011, 2011, pp. 9–15.
- [15] B. Taspinar, W. Schmidt, and H. Schuhbauer, "Gamification in education: A board game approach to knowledge acquisition," in *Procedia Computer Science*, 2016, vol. 99, pp. 101–116.
- [16] B. E. Wiggins, "An overview and study on the use of games, simulations, and gamification in higher education," *Int. J. Game-Based Learn.*, vol. 6, no. 1, pp. 18–29, 2016.
- [17] S. A. Kocadere and Ş. Çağlar, "The design and implementation of a gamified assessment," *J. E-Learning Knowl. Soc.*, vol. 11, no. 3, pp. 85–99, 2015.
- [18] L. Moccozet, C. Tardy, W. Opprecht, and M. Leonard, "Gamification-based assessment of group work," in *2013 International Conference on Interactive Collaborative Learning, ICL 2013*, 2013, pp. 171–179.
- [19] Z. Zainuddin, M. Shujahat, H. Haruna, and S. K. W. Chu, "The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system," *Comput. Educ.*, vol. 145, 2020.
- [20] Y. Attali and M. Arieli-Attali, "Gamification in assessment: Do points affect test performance?," *Comput. Educ.*, vol. 83, pp. 57–63, 2015.
- [21] M. Guerreiro and C. Nordengren, "'No Fun Games': Engagement Effects of Two Gameful Assessment Prototypes," *J. Res. Technol. Educ.*, vol. 50, no. 2, pp. 134–148, 2018.
- [22] K. Berkling and C. Thomas, "Gamification of a software engineering course and a detailed analysis of the factors that lead to its failure," in *2013 International Conference on Interactive Collaborative Learning, ICL 2013*, 2013, pp. 525–530.
- [23] D. Rojas, B. Kapralos, and A. Dubrowski, "The missing piece in the gamification puzzle," in *ACM International Conference Proceeding Series*, 2013, pp. 135–138.
- [24] R. N. Landers, "Gamification Misunderstood: How Badly Executed and Rhetorical Gamification Obscures Its Transformative Potential," *J. Manag. Inq.*, vol. 28, no. 2, pp. 137–140, 2019.
- [25] 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," *Comput. Human Behav.*, vol. 69, pp. 371–380, 2017.
- [26] R. van Roy and B. Zaman, "Why gamification fails in education and how to make it successful: Introducing nine gamification heuristics based on self-determination theory," in *Serious Games and Edutainment Applications: Volume II*, 2017, pp. 485–509.
- [27] L. Shi and A. I. Cristea, "Motivational gamification strategies rooted in self-determination theory for social adaptive e-learning," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9684, pp. 294–300.
- [28] S. Nicholson, "A recipe for meaningful gamification," in *Gamification in Education and Business*, Springer, Cham, 2015, pp. 1–20.
- [29] A. F. Aparicio, F. L. G. Vela, J. L. G. Sánchez, and J. L. I. Montes, "Analysis and application of gamification," in *INTERACCION '12: Proceedings of the 13th International Conference on Interacción Persona-Ordenador*, 2012, pp. 1–2.
- [30] D. H. Rose, A. Meyer, N. Strangman, and G. Rappolt, *Teaching every student in the digital age: universal design for learning*. Alexandria, VA: Association for Supervision & Curriculum Development (ASCD), 2002.
- [31] M. Böckle, I. Micheel, M. Bick, and J. Novak, "A Design Framework for Adaptive Gamification Applications," in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
- [32] N. Gordon, M. Brayshaw, and S. Grey, "Maximising gain for minimal pain: Utilising natural game mechanics," *Ital. Innov. Teach. Learn. Inf. Comput. Sci.*, vol. 12, no. 1, pp. 27–38, 2013.
- [33] T. Jagušt, I. Botički, and H. J. So, "Examining competitive, collaborative and adaptive gamification in young learners' math learning," *Comput. Educ.*, vol. 125, pp. 444–457, 2018.
- [34] A. Iosup and D. Epema, "An experience report on using gamification in technical higher education," in *SIGCSE 2014 - Proceedings of the 45th ACM Technical Symposium on Computer Science Education*, 2014, pp. 27–32.
- [35] S. J. Aguilar, C. Holman, and B. J. Fishman, "Game-Inspired Design: Empirical Evidence in Support of Gameful Learning Environments," *Games Cult.*, vol. 13, no. 1, pp. 44–70, 2018.
- [36] P. Sweetser and P. Wyeth, "GameFlow: a model for evaluating player enjoyment in games," *Comput. Entertain.*, vol. 3, no. 3, 2005.
- [37] A. S. Miller, J. A. Cafazzo, and E. Seto, "A game plan: Gamification design principles in mHealth applications for chronic disease management," *Health Informatics J.*, vol. 22, no. 2, pp. 184–193, 2014.

- [38] F. L. Fu, R. C. Su, and S. C. Yu, "EGameFlow: A scale to measure learners' enjoyment of e-learning games," *Comput. Educ.*, vol. 52, no. 1, pp. 101–112, 2009.
- [39] J. Hamari and J. Koivisto, "Measuring flow in gamification: Dispositional Flow Scale-2," *Comput. Human Behav.*, vol. 40, pp. 133–143, 2014.
- [40] H. Elmunsyah, W. N. Hidayat, and K. Asfani, "Interactive learning media innovation: utilization of augmented reality and pop-up book to improve user's learning autonomy," *J. Phys. Conf. Ser.*, vol. 1193, no. 1, 2019.