International Journal of Serious Games

ISSN: 238
4-8766
https://journal.seriousgamessociety.org/

Article

A Bird Matching Game: Difficulty, Rewards, and Intrinsic Motivation

Janne Tyni[1], Roman Bednarik[1], Juho Kahila[2] and Matti Tedre[1]

[1]*School of Computing, University of Eastern Finland, Joensuu, Finland;* [2]*School of Applied Educational Science and Teacher Education, University of Eastern Finland, Joensuu, Finland*
janne.tyni@uef.fi;roman.bednarik@uef.fi;juho.kahila@uef.fi;matti.tedre@uef.fi

Keywords:

Digital Game Based Learning
Engagement
Extrinsic motivation
Gamification
Game mechanics
Intrinsic motivation
Motivation
Serious games

Received: May 2024
Accepted: October 2024
Published: November 2024
DOI: 10.17083/ijsg.v11i4.805

Abstract

This study investigates the effects of optional difficulty settings, badges, points, and educational rewards on intrinsic motivation. We provide insights into how reward structures affect intrinsic motivation and in-game progression, tested in a bird-matching game where players match species with names. These findings are significant for designing educational games that effectively engage and motivate players. We created two versions of the game: one version with 'pointification' rewards (badges and points) and another with educational rewards (learning materials and bird sounds). The effect of voluntarily increasing the difficulty of the game is also investigated. In-game analytics of 66 sessions are examined. In addition, the players' Intrinsic Motivation Inventory (IMI) survey scores are calculated. Further statistical analyses (t-tests, visualizations and other calculations) were performed using R software. The analyses indicate that 'pointification' and educational rewards show no statistically significant difference in either the in-game progression or IMI scores of players. The findings show that intrinsic motivation is unaffected by either points and badges or educational rewards. However, voluntarily increasing the difficulty is shown to increase the IMI scores of players. This study expands on the discussion of alternatives to 'pointification' and provides new insights on the optional difficulty systems in educational games.

1. Introduction

Researchers who study serious games often read publications such as [3], which discuss a field of research positioned between utilitarian systems (such as a banking system, used for a purpose) and hedonic systems (like games, used for enjoyment), known as "gamification". There is no one universally accepted definition for this term. One of the most often cited is "the use of game design elements in non-game contexts" [1]. The big idea is to "transfer the

motivational effects of games into education", as [4] put it. The question of where and how gamification should be applied (and to what effect) is still left unanswered [2].

The term "gamification" was documented for the first time in 2008 [1]. The number of publications in this field of research has rapidly increased after the year 2010 [1,5]. The findings about the effects of gamification on motivation and engagement are inconsistent [2], and there is no consensus on when and where to apply gamification [2].
In this study we investigate the effects of optional difficulty settings, badges, points and educational rewards on intrinsic motivation. This is done in the context of an educational bird-matching game where players match bird species with names. We created two different versions of the game, one featuring 'pointification' rewards such as points and badges and the other offering educational rewards such as additional learning materials and bird sounds. This study examines the in-game analytics collected from 66 gameplay sessions in addition to evaluating the results of a questionnaire that is designed to evaluate intrinsic motivation.

The following sections will discuss the current state of gamification research, including various gamification elements, the role of intrinsic and extrinsic motivation, methods of measurement, and current debates in the literature. Furthermore, we will identify the gaps in literature and explore the use of alternative rewards and difficulty settings in motivation research.

1.1 Background

The field of gamification research is ripe with studies about badges and points [6]. Some studies say that adding points, achievements and badges to your project will boost motivation and increase self-efficacy and that they are "loved by most students" [7]. Indeed, a 2020 study that surveyed the perception of gamification elements from students found that levels and points were among the most highly valued gamification elements while timer a was the least valued [8].

Even if end users love points and badges, how much reward is enough and in what context? Short answer: we do not know. Recreational games have more variety in different types of rewards when compared to educational games [9]. Studies have found that the greater variety of rewards could positively influence interest and enjoyment [10] and provide a better experience [11]. The sense of effort and enjoyment can even be enhanced by giving out greater rewards [11]. Others, on the other hand, say that the number of in-game rewards does not affect the sense of feeling rewarded [12]. The body of research shows mixed results.

If our understanding of gamification and its proper application were to improve, we could increase engagement metrics. Individual studies have shown that gamification can be used to increase participation in online courses [13] and increase the time that a system is used for [14]. It has been stated that "motivated engagement is essential in educational interventions" and that the design of an educational game that produces such engagement is not an easy feat [15].

Some say that we should focus on the core mechanics and making the game more enjoyable to the player rather than fixating on in game rewards [12]. For example, the difficulty of a game has been tested and shown that self-selected difficulty is the most motivating [16]. If randomly given difficulty, the easiest difficulty setting appears to motivate the most [16]. In short, there are multiple different ways to implement gamification and many of them seem to affect motivation. But this raises more questions: what exactly is motivation and how do we measure the effectiveness of an individual game mechanic or a gamification element?

To understand what motivation is, we must discuss the two kinds of motivation that are often investigated in gamification research: Intrinsic and Extrinsic motivation. Intrinsic motivation is generally defined as the internal drive to take an action rather than doing it for an outside (extrinsic) reward [17]. In practice the difference is studying bird names because you want to learn to recognize species in the wild (intrinsic motivation) and studying bird names because you want to get a good score from an exam (extrinsic motivation).

The effects of gamification elements on intrinsic motivation have been a research subject for a while now. A systematic review of studies between 1990 and 2020 shows that badges and points may improve intrinsic motivation [18], while some empirical studies report that intrinsic motivation is not affected by badges [19, 20]. Now that we understand what motivation is (and that we have no clue what element affects it and why), we can delve into how it's measured.

The Intrinsic Motivation Inventory (IMI) survey is one of the most frequently used standardized questionnaires in games user research [21]. It is based on self-determination theory, which is popular in gamification studies [22]. The original survey was first used in 1982 [23] and the current iteration has six subscale scores: interest/enjoyment, perceived competence, effort, value/usefulness, felt pressure and tension in addition to perceived choice while performing a given activity [24]. The subscales measure their namesakes and are calculated based on several questions such as "This activity was fun to do." That is rated on a scale of 1 (not at all true) to 7 (very true). The first subscale, interest/enjoyment is the most used subscale in games user research [25]. It is also the only subscale in the survey that specifically assesses intrinsic motivation [24].

The use of standardized questionnaires makes the comparison between studies easier. The easier the comparison between the studies, the easier it should be to identify which gamification elements have effects on intrinsic motivation. Despite this obvious advantage that comes with using standardized questionnaires, the number of studies in gamification that use these to measure outcomes is low [21].

The IMI survey has been used in a wide variety of fields and applications. For example, [26] used the IMI survey to measure the effect of a serious game on children undergoing urotherapy, but no significant effects were found. [27] studied shallow gamification elements in a mathematics trails application and found that, while performance parameters improved, there was no significant impact on motivation. [28] used IMI to study motivation in a priorisation game. They measured on the subscales of value/usefulness, enjoyment/interest and effort/importance. They found a statistically significant increase in intrinsic motivation and enjoyment with the gamified task. [29] used IMI interest/enjoyment, effort, competence/autonomy subscales to study points, leaderboards and badges. They found that badges and a leaderboard contribute to intrinsic motivation and performance.

1.2 Knowledge gap

Gamification is a widespread practice but only a few studies are carefully designed and empirically tested [30]. The mixed results in the literature have left us, gamification researchers, confused about which gamification elements should be implemented where and why [2]. Our understanding of the influence of gamification on our motivation and behavior has been discussed as a "black box" [32]. Most of the studies that discuss rewards in gamification are discussing theory and there is a need for practical studies [5].

The empirical research should include "surveys that measure latent psychological variables" [13]. The goals of the individual users should be considered [3] and the studies should be directed to discovering alternatives to points and badges [32, 33].

Some studies have been done to expand the concept of reward to educational content. A week-long study that focused on sustainability marketing outcomes found that badges and trophies significantly enhanced knowledge in the field. Comparatively, the same study found that educational rewards had little impact [34]. Educational cards were used as a reward for fifth grade students, and it was found that the application did motivate the students to use a vocabulary learning system [35]. There is a dire need to expand the literature that explores alternatives like this and measures their outcomes with standardized surveys.

There are studies that implicate that gamification can have a positive effect on intrinsic need satisfaction [32] but the effect of individual gamification elements on intrinsic motivation remains a mystery [29] The gamification studies in the context of educational games appear promising but the influence of gamification on extrinsic and intrinsic motivation is left unexplored [36].

In short, we do not know which gamification elements affect what outcomes. Reviews on the field comment frequently on the lack of studies that systematically investigate this [3, 32]. To answer these calls to empirical research, we must procure a variety of different gamification elements to test out and measure the effects using standardized questionnaires.

1.3 Objectives of the study

The mystery behind the "black box" of the effects of individual gamification elements on intrinsic motivation has us excited. There is much to learn and clear-cut ways to achieve knowledge. In this study we will expand on the body of literature in gamification and reward research in new directions. This will be done by altering a simple bird name matching game. The educational game consists of matching a picture of a bird to its correct name.

We will create two versions of the game. One version will reward players with "pointification" rewards (points and badges, as described by [4]) and customizable background options. The second version will explore alternative options to pointification rewards. These rewards will be educational and perhaps more valuable to the players who have intrinsic motivation to learn bird names. The rewards we have chosen for the second version are: bird names in latin, bird sounds and additional information about the birds.

In addition to exploring the effect of these educational rewards when compared to "pointification" rewards, we will explore the effect of using optional difficulty modifiers. The students will have the opportunity to voluntarily add a timer, increase the number of correct guesses or the answer options and change the language of the bird names to latin. Interested students will be randomly given the "pointification" or the "educational" version of the bird matching game (with the option to use difficulties if they so choose) online, along with the IMI survey.

By comparing the IMI questionnaire responses between the "pointification" and "educational" versions and the students who used or did not use optional difficulty modifiers we hope to discover answers to the following research questions:

1. How do the IMI questionnaire scores of the students differ with the “pointification” rewards when compared to “educational” rewards?
2. How do the IMI questionnaire scores of the students differ with the use of optional difficulties when compared to those who did not use difficulties?
3. Do the “pointification” rewards, “educational” rewards or the use of optional difficulty modifiers affect how far the players progress in the game?

2. Methodology

Data for the study was collected during the year 2024 from the January 29th to the 5th of February. During this time the biology/geography departments of four large universities of Finland were contacted with the opportunity to participate in the research. The students were shared an information leaflet in their respective information channels. The leaflet instructed the players to participate in the research by playing the bird naming quiz game online. Individuals who had connections to bird watching groups contacted the researchers and asked for a permission to share the leaflet to other interested participants and the permission for this was granted. The players were able to participate in the research by playing the browser-based game on their computers or other mobile devices. The data regarding decisions made during the game and the Intrinsic Motivation Inventory survey [23] was sent to the research team after a play session and was analyzed.

Questions from the IMI questionnaire sections "interest/enjoyment" (Table 1), "effort/importance" (Table 2) and "value/usefulness" (Table 3) were chosen as they suit the purpose of answering the research questions of this study. The questions were randomly ordered and presented to the players in the end survey available at the end of the play session in game.

Table 1. IMI questionnaire questions for interest/enjoyment. Seven items in total.

Number	Likert statement (7-point scale)
1	I enjoyed doing this activity very much.
2	This activity was fun to do.
3	I thought this was a boring activity.
4	This activity did not hold my attention at all.
5	I would describe this activity as very interesting.
6	I thought this activity was quite enjoyable.
7	While I was doing this activity, I was thinking about how much I enjoyed it.

Table 2. IMI questionnaire questions for effort/importance. Five items in total.

Number	Likert statement (7-point scale)
1	I put a lot of effort into this.
2	I didn't try very hard to do well at this activity.

3	I tried very hard on this activity.
4	It was important to me to do well at this task.
5	I didn't put much energy into this.

Table 3. IMI questionnaire questions for the average value/usefulness. Seven items in total.

Number	Likert statement (7-point scale)
1	I believe this activity could be of some use to me.
2	I think that doing this activity is useful for learning bird names.
3	I think this is important to do because it can teach bird names.
4	I would be willing to do this again because it has some value to me.
5	I think doing this activity could help me to learn bird names.
6	I believe doing this activity could be beneficial to me.
7	I think this is an important activity.

The data was split into different categories. First, the players with the “educational” (edu) and “pointification” (point) versions of the game were separated and analyzed. In a similar manner the players who used the difficulty system were categorized into two data sets: the UD (Used Difficulties) and the DUD (Did not Use Difficulties). Further analysis (T tests, visualizations and other calculations) was done using the R-software (version 4.1.3).

2.1 The game

Before the game is first loaded the game randomly assigns the player with either the educational reward version or the point version of the game. When the game is first loaded, the player is asked to select the language; the given options are English and Finnish. After reading about the purpose of the research the player is presented with the level map (Figure 1). The level map is designed in a way that allows the player to choose the levels they wish to play. At the very beginning, two level options are available (as can be seen from Figure 1).

Figure 1. The level map is the first screen that the player is presented at the beginning of the game. In the middle the levels of the game are presented. Skull icons with orange coloring represent available levels. Lock icons with a grey background represent levels that are currently locked. There are buttons for settings and information on the bottom left and top right corners of the screen, respectively.

From the level map the player can access and view different information screens that are available for each level (Figure 2). The information screens show information about the rewards presented in the game and the individual birds. The birds that are to be named in this game have been selected from common Finnish birds that appear in university level courses. The game levels contain different bird families that have been grouped together. For example, one of the first levels contains birds from the Gaviiformes, Pelecaniformes and Anseriformes families (Figure 2).

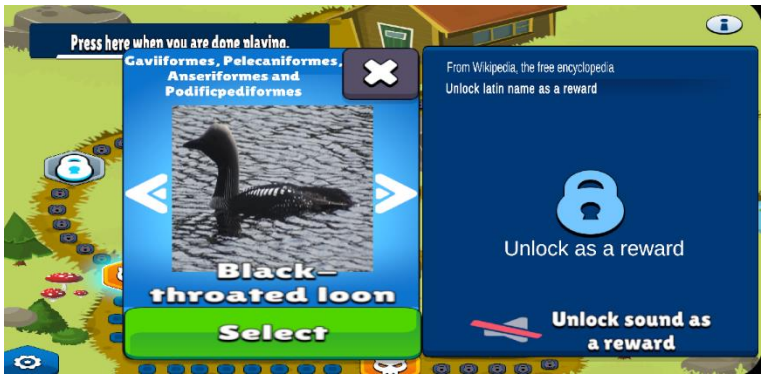

Figure 2. Info screen of a level in the game. On the left the bird families that this level consist of can be seen. Below that the individual bird pictures and their names can be viewed by pressing the buttons on the sides of the image of the bird. Another window shows some of the reward types that can be unlocked with gameplay.

After selecting the level from the info screen the player is presented with the difficulty screen (Figure 3). The difficulty screen contains further details about the available difficulties and change when difficulties are added or removed.

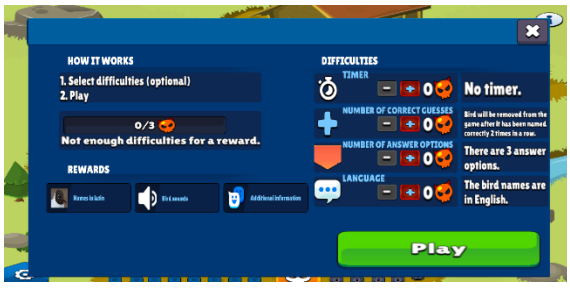

Figure 3. The difficulty screen. This is a screen that is shown right before gameplay and allows the player to select additional difficulties. Difficulties with explanations are on the right side of the screen. Instructions of use are in the top left corner. The bar below instructions fills or depletes when difficulties are either added (+ -buttons) or removed (- -buttons).

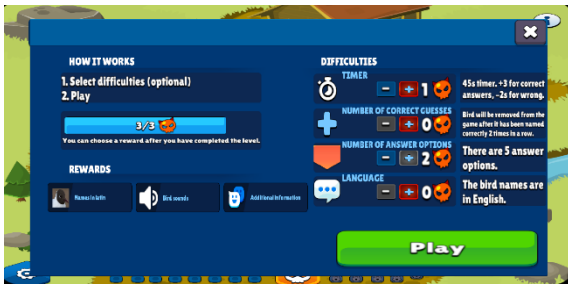

Figure 4. The difficulty screen with three difficulties selected. The selected difficulties are 1 count of timer and two counts of additional answer options. The bar on the left has been filled and an information text under it states that a reward can be chosen after the level has been completed.

The difficulties that are available are a timer, an increase of number of correct guesses, an increase in the number of answer options and a change in the language in which the bird names are. Table 4 shows a more detailed list of the available difficulties.

Table 4. A detailed table of the difficulties available. Each difficulty has multiple levels ranging from 0 (no difficulty) to difficulty level 3 (most difficult). The timer always gives 3 seconds more to the timer for correct answers and reduces 2 seconds from wrong answers. The difficulty level 0 language of the game depends on the language that the player chose at the beginning of the game session.

Difficulty	Level 0	Level 1	Level 2	Level 3
Timer	No timer	45 second timer	30 second timer	15 second timer
Number of correct guesses	2	3	4	5
Number of answer options	3	4	5	-
Language	English/Finnish	Latin	-	-

Once three difficulties are added, the screen explains that conditions have been met for a reward after successfully completing the level, as can be seen from Figure 4. The game can be started with or without difficulties by pressing the "Play" button at the bottom right of the screen.

The gameplay is designed to work as a quiz. Each level contains a variety of bird images from the selected level. The player is presented with an image of the bird and several buttons with different bird names (Figure 5). Each bird has multiple different pictures that appear in random order and have to be named correctly a set time in a row. In the Figure 5 it can be seen from the top left that the difficulty selected for this level is set to two times in a row. After being correctly named, the player is then presented with another bird from the pool of 27 possible birds on this level (Figure 5). Once the bird has been named correctly a set number of times, the pool of possible birds shrinks from 27 to 26 and so on until all the birds have been named. If the player is not successful in the attempt and there is a difficulty in place that enables failing the level (the timer) the player is presented with the option to either go back to the level map screen or to restart the level. If the player is successful at naming all the birds the player is presented with a stage clear screen (Figure 6). If the player has selected enough difficulties, there will be different rewards to select from depending on the version of the game that the player had access to. If the player was assigned the point version of the game this screen also presents a score that is calculated based on the number of birds correctly named.

Figure 5. The gameplay screen. In the middle of the screen there can be seen a picture of a Black-throated loon. The player has to press the correct name from the five different bird name options on the bottom of the screen. A timer can be seen running down on the left side of the screen. The bird has been correctly named 0 out of the required 2 times as can be seen from the top left of the screen. An exit button is at the bottom left of the screen.

Figure 6. The stage clear screen. This screen presents the player with an option to select one of three educational rewards.

Before the play session started the game assigned the player randomly with either educational version or the point version of the game. The only difference between these two versions are the rewards available. Table 5 shows a detailed list of the different rewards included in each version.

Table 5. Table with the list of different rewards available in each version of the game. The rewards are in no particular order. The rewards are accessible from the info screen that can be seen before selecting a level on the level map screen.

Version	Reward 1	Reward 2	Reward 3
Educational (Edu)	Bird names in latin	Bird sounds	Additional information about the birds
Pointification (Point)	Badge	Double score	Background options

After selecting a reward or ending the level the player is presented with the level screen map (Figure 1). If the level was successfully completed the map has changed to open new levels for play (As can be seen in Figure 7). After enough levels with the difficulties have been completed and the rewards unlocked the info screen of the player shows the earned rewards (Figure 8).

Figure 7. The level map screen. In this screen a level has been successfully completed (as indicated by the blue color and the lack of a skull icon). Routes to new levels have been unlocked.

Figure 8. The info screen of a player that has unlocked all of the “pointification” (point) version rewards.

After the player decides that they have played enough of the game, the level map screen (Figures 1 and 7) shows a button stating, "Press here when you are done playing.". This button opens the IMI questionnaire that can be filled and sent to the research team. This ends the game session.

3. Results

During the data collection period 66 unique play sessions were gathered for further analysis. 31 play sessions were collected from players that were assigned the “educational” (edu) version of the game. The remaining 35 sessions were collected from players of the “pointification” (point) version. The data collected from these two versions were separated and analyzed. In addition, the analysis separates the 31 participants who used the difficulty system, UD (Used Difficulties), from the 35 players who opted not to add difficulties, DUD (Did not Use Difficulties), to their game session. The analysis between the two different versions and the difficulty user analysis are done separately from each other.

3.1 Educational and pointification data sets

35 Participants played the version of the game with educational rewards. 31 Participants played the version where the rewards were double points, a badge and customized backgrounds. The edu and point versions had no statistically significant difference in the number of levels completed in the data sets ($t = 0.21564$, $df = 64$, p-value $= 0.83$). On average the Edu data set participants completed 3.4 levels and the Point data set participants completed 3.2 levels. The mean levels completed was 3.3 for the edu version and 3.1 for point version. Figure 9 shows that both versions had players who didn't complete any levels and players that completed all of them.

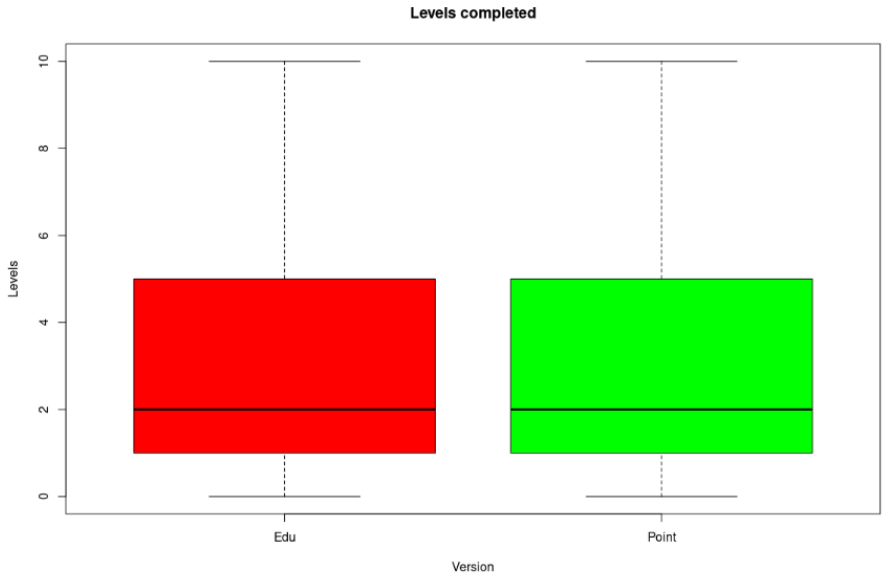

Figure 9. The variance and mean of the number of levels completed for the “Educational” (Edu) and “Pointification” (Point) versions of the game.

Analysis for the IMI survey responses for the average interest/enjoyment in both data sets show that there is no statistically significant difference in any of the seven questions that were asked from the players (Table 6). The biggest difference in these average scores is 0.2. The variance in average score is slightly larger in the point data set (Figure 10).

Table 6. IMI questionnaire responses for the average interest/enjoyment. Edu is the educational rewards data set and Point is for the gamified rewards data set. Total score is calculated by summing the statements. The score for the statements with (R) behind them is calculated by first subtracting 8 from the score. This is done because the statements have a negative connotation when compared to the other questions.

	Likert statement (7-point scale)	Edu	Point	p
1	I enjoyed doing this activity very much.	4.3	4.4	0.8725
2	This activity was fun to do.	4.8	4.8	0.9673
3	I thought this was a boring activity. (R)	2.5	2.3	0.5284
4	This activity did not hold my attention at all. (R)	2.0	2.1	0.6915
5	I would describe this activity as very interesting.	4.0	4.1	0.8972
6	I thought this activity was quite enjoyable.	4.2	4.2	0.9335
7	While I was doing this activity, I was thinking about how much I enjoyed it.	3.7	3.6	0.8995
	Total score	32.5	32.7	0.9578

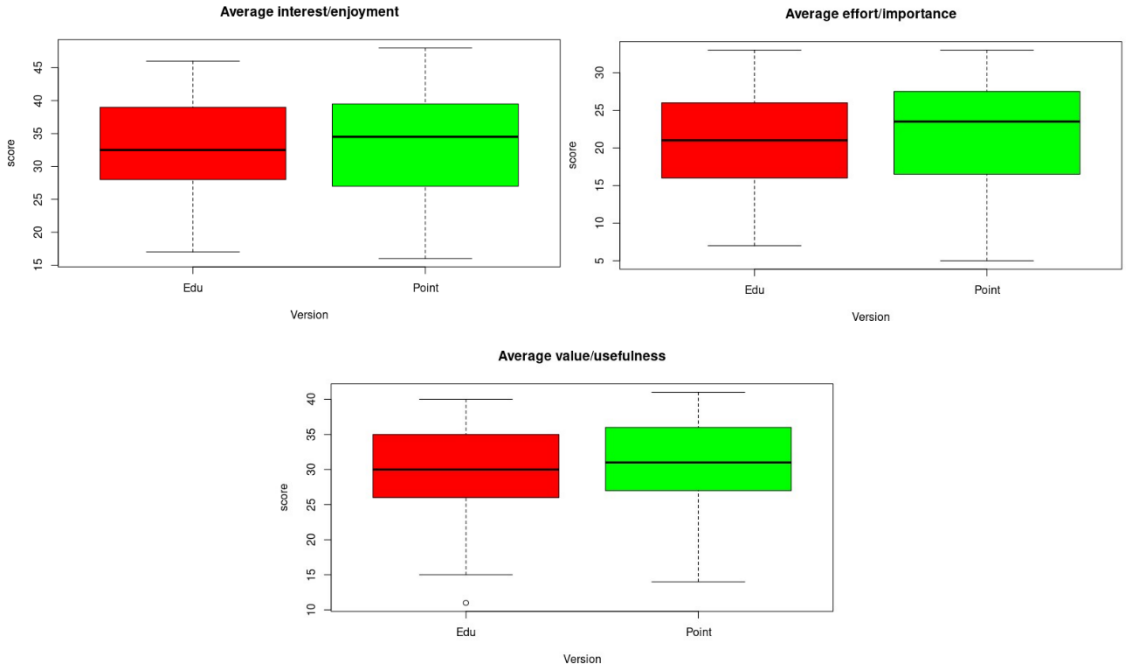

Figure 10. The mean interest /enjoyment score for participants who had the educational version of the game was 32.50 while the participants who had the point version of the game had a mean score of 32.61. The mean for effort/importance was 20.50 for the edu version and 21.7 for the point version. The average value/usefulness score for the edu version is 29.73 while the point version had 30.64.

The IMI survey responses for effort/importance analysis shows that there are no statistically significant differences in any of the questions asked (Table 7). The table shows that there is a 1.2 score difference for the average total scores for both data sets. The point data set had a slightly larger variance and mean as can be seen from Figure 10.

Table 7. The IMI questionnaire responses for the average effort/importance. Edu is the educational rewards data set and Point is for the gamified rewards data set. Total score is calculated by summing the statements. The score for the statements with (R) behind them is calculated by first subtracting 8 from the score.

	Likert statement (7-point scale)	Edu	Point	p
1	I put a lot of effort into this.	3.8	3.9	0.7754
2	I didn't try very hard to do well at this activity. (R)	3.5	3.2	0.5538
3	I tried very hard on this activity.	4.5	4.8	0.5958
4	It was important to me to do well at this task.	3.9	4.3	0.3808
5	I didn't put much energy into this. (R)	4.2	4.1	0.6617
	Total score	20.5	21.7	0.5007

Similarly, the IMI average responses for value/usefulness show no statistically significant difference (Table 8) and the point version has a slightly larger mean (Figure 10).

Table 8. The IMI questionnaire responses for the average value/usefulness. Edu is the educational rewards data set and Point is for the gamified rewards data set. Total score is calculated by summing the statements.

	Likert statement (7-point scale)	Edu	Point	p
1	I believe this activity could be of some use to me.	5.0	5.4	0.2801
2	I think that doing this activity is useful for learning bird names.	5.6	5.5	0.8365
3	I think this is important to do because it can teach bird names.	4.9	5.4	0.1776
4	I would be willing to do this again because it has some value to me.	4.6	4.3	0.6182
5	I think doing this activity could help me to learn bird names.	5.6	5.8	0.4961
6	I believe doing this activity could be beneficial to me.	5.0	5.0	0.873
7	I think this is an important activity.	4.0	4.6	0.148
	Total score	34.7	36.0	0.599

3.2 Difficulty data sets

On average the "Used Difficulties" (UD) data set participants completed 3.3 levels and the "Didn't Use Difficulties) (DUD) data set participants completed 3.1. The use of different difficulty options spread evenly between the Edu and Point data sets (Table 9). The difference in the mean levels completed for the UD and DUD data sets was 0.11 (Figure 11). 35 Participants chose not to increase the difficulty of gameplay. The number of participants who used at least one difficulty is 31. Out of the participants who increased the difficulty 15 of them chose to take a reward at the end of the game session. The analysis shows that there is no statistically significant difference in the number of levels completed in either UD or DUD data sets ($t = -0.19782$, $df = 64$, p-value = 0.8438).

Table 9. A detailed table of the number of participants who increased the difficulty of gameplay and the difficulties they used. For example, 10 participants from the educational data set used a timer. The "educational" (edu) and "pointification" (point) data sets are represented.

Data set	Timer	Number of correct guesses	Number of answer options	Latin
Edu	10	11	10	5
Point	12	13	17	5
Total number of participants	22	24	27	10

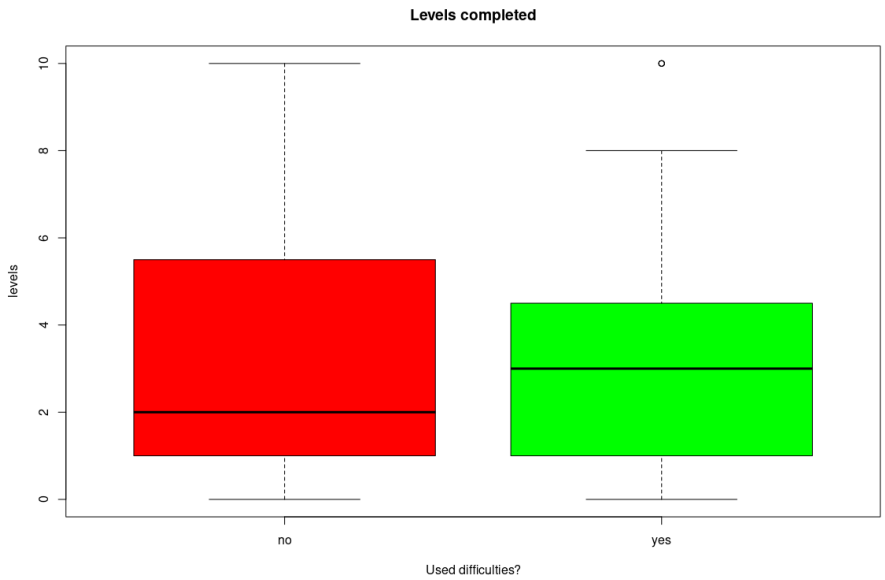

Figure 11. The variance and mean for levels completed for "educational" (edu) and "pointification" (point) data sets.

The analysis for the IMI survey responses for the average interest/enjoyment shows that the fourth question (This activity did not hold my attention at all) was statistically significantly different between the data sets (Table 10). The data set as a whole had an average score of 23.5 for the UD data set and 19.1 for the DUD data set. Figure 12 shows that the participants who used difficulties showed a slightly smaller variance in responses.

Table 10. The IMI questionnaire responses for the average interest/enjoyment. UD (Used Difficulties) participated by raising the difficulty of their game and DUD (Didn't Use Difficulties) did not. Total score is calculated by summing the statements. The score for the statements with (R) behind them is calculated by first subtracting 8 from the score. p-values less than 0.05 are marked with a *.

	Likert statement (7-point scale)	UD	DUD	p
1	I enjoyed doing this activity very much.	4.3	4.4	0.7085
2	This activity was fun to do.	4.8	4.8	0.9385
3	I thought this was a boring activity. (R)	2.3	2.5	0.5174
4	This activity did not hold my attention at all. (R)	1.7	2.3	0.03612*
5	I would describe this activity as very interesting.	4.1	4.0	0.8598

6	I thought this activity was quite enjoyable.	4.3	4.1	0.7181
7	While I was doing this activity, I was thinking about how much I enjoyed it.	3.7	3.6	0.7528
	Total score	23.5	19.1	0.6085

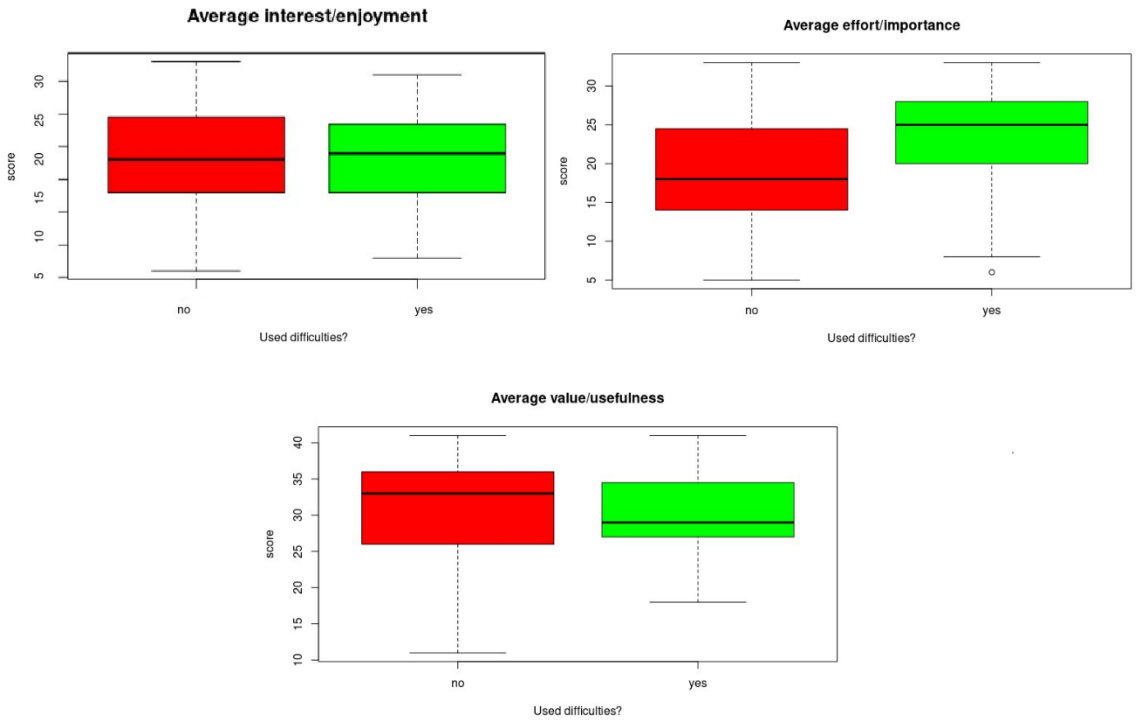

Figure 12. Three boxplots showing the variance and mean for the measured IMI survey results. The means of average interest/enjoyment, effort/importance and value/usefulness. All the boxplots showcase the differences between the UD and DUD data sets.

Table 11 shows that the comparison between UD and DUD data sets in the average effort/importance analysis shows several statistically significantly different average scores. The questions 3-5 and the total score all show significant difference. These items were "I tried very hard on this activity", "It was important to me to do well at this task" and "I didn't put much effort into this". The total average score for the UD data set was 23.5 while the DUD average total score was 19.1. The mean in the data set who used difficulties is higher, as can be seen in Figure 12.

Table 11. The IMI questionnaire responses for the average effort/importance. UD (Used Difficulties) participated by raising the difficulty of their game and DUD (Didn't Use Difficulties) did not. Total score is calculated by summing the statements. The score for the statements with (R) behind them is calculated by first subtracting 8 from the score. p-values less than 0.05 are marked with a *.

	Likert statement (7-point scale)	UD	DUD	p
1	I put a lot of effort into this.	4.2	3.6	0.1657
2	I didn't try very hard to do well at this activity. (R)	3.0	3.6	0.1885
3	I tried very hard on this activity.	5.2	4.1	0.01569*
4	It was important to me to do well at this task.	4.6	3.7	0.0238*
5	I didn't put much energy into this. (R)	3.5	4.7	0.004811*
	Total score	23.5	19.1	0.01215*

The analysis of the IMI survey responses for the value/usefulness (Table 12) reveal no statistically significantly different results from either data set. The average total score differs by 0,6. The variance and mean in this data set is slightly higher in the DUD data set (Figure 12).

Table 12. The IMI questionnaire responses for the average value/usefulness. UD (Used Difficulties) participated by raising the difficulty of their game and DUD (Didn't Use Difficulties) did not. Total score is calculated by summing the statements.

	Likert statement (7-point scale)	UD	DUD	p
1	I believe this activity could be of some use to me.	5.4	5.1	0.5507
2	I think that doing this activity is useful for learning bird names.	5.5	5.5	0.9339
3	I think this is important to do because it can teach bird names.	5.1	5.2	0.713
4	I would be willing to do this again because it has some value to me.	4.1	4.7	0.2079
5	I think doing this activity could help me to learn bird names.	5.8	5.5	0.5174
6	I believe doing this activity could be beneficial to me.	4.9	5.1	0.523
7	I think this is an important activity.	4.2	4.5	0.3949
	Total score	35.0	35.6	0.6959

4. Discussion

4.1 Main Results

In this research we set out to figure out whether or not the different aspects of intrinsic motivation would differ between players if they were given the “pointification” (point) or “educational” (edu) versions of the game (RQ1). It is interesting to note that the differences between the edu and point data sets were only slight in the IMI surveys. In terms of the average interest/enjoyment, the Table 6 shows that there were no statistically significant difference in the average statement scores. The total score had only a 0.2 difference between the two groups. Similarly, the average effort/importance scores for these two groups show only marginal differences as the total average score differs only 1.2 points between the groups (Table 7). The largest difference between the edu and point versions was found to be the 1.3 point total score difference in the value/usefulness, but even these scores were found not to be statistically significantly different from each other (Table 8). The means and variance in all of the metrics show similarities (Figure 10). With these results in mind, it can be said that to answer the first research question, there is no difference in intrinsic motivation of players that got either the edu or point version of the game.

The second question this paper set out to answer was if the use of in game difficulties affected the intrinsic motivation aspects of the players (RQ2). It is interesting to note that the number of participants who used different difficulties and the spread of difficulty use is fairly similar in both versions of the game (Table 9). The UD (Used Difficulties) and DUD (Didn't Use Difficulties) groups had statistically significantly different average answers to the intrinsic motivation inventory questionnaire subset “effort/importance”, but not in other subsets. In the interest/enjoyment questions the players had a difference in opinion where the DUD group had scored the statement 4 "This activity did not hold my attention at all." significantly more when

compared to the UD group (Table 10). Similarly, the variance and mean of these two groups in the average interest/enjoyment questions are similar, though the UD group has a slightly smaller variance (Figure 12). The most difference that can be found between the UD and DUD group answers is in the average effort/importance Table 11 where the statements 3-5 (I tried very hard on this activity.", "It was important for me to do well at this task." and "I didn't put much energy into this.") all show statistically significantly different answers. The total average score for the whole effort/importance is also significantly different with scores 23.5 for the UD group and 19.1 to the DUD group. Similarly, the mean is greater for the UD group when compared to the DUD group in Figure 12. The value/usefulness answers show no statistically significant differences between the two groups in Table 12, and the total scores of the two groups differ by 0.6 points. The variance for the DUD group is noticeably larger in the average value/usefulness scores as can be seen in the Figure 12. In conclusion, it can be said that the use of in game difficulties significantly affected the perceived effort/importance answers to the IMI questionnaire.

The third question that this paper wanted to explore was whether or not the version or the usage of optional difficulties correlated with the in-game progression of players (RQ3). Figure 9 shows that both the edu and point versions of the game had roughly the same variance. The average number of levels progressed in the game for the edu group was 3.4 and 3.2 for the point data set, showing little difference. In regard to the UD and DUD groups, Figure 10 shows that there is greater variance. The group who used difficulties in their gameplay completed on average 3.3 levels where the DUD group who didn't use difficulties completed 3.1. In regard to the third research question, it can be said that, as all of the average levels completed were between 3.1 and 3.4, the usage of difficulties or version of the game made no difference to the progress the players made inside the game.

This study has several implications. First, this study shows that points and badges appear to be as effective in altering the IMI questionnaire results as are educational rewards. This suggests that more research is needed in the field of rewards when it comes to alternatives for points and badges to further understand all the possible alternatives and the use cases. Second, the players who did use difficulties showed significantly different results in the effort/importance scale of IMI survey. This suggests that the potential in optional difficulties needs further exploring in the field of gamification. Third, the use of difficulties or the version of game that the players got did not seem to affect the progression of players at all. How to motivate players to advance further in the game is a field that requires more rigorous empirical study.

4.2 Limitations

In this study there are multiple limitations. The study was done with Finnish participants and was not a longitudinal study. This study consists of only 66 in-game sessions. It is possible that with more sessions, the results of analyses may change. Similarly, measuring the intrinsic motivation of the participants before participating in the activity may reveal additional insights. The participants were not carefully monitored. As the participants had the freedom to play the game with any browser-based system of their choosing, it is possible that some of the systems may have acted in a way different than intended. If malfunctions occurred, it could have affected the collected metrics. Background information from students was not collected, so the demographics, prior experience with games and familiarity with the topic of the game (bird species) was not considered when performing analyses. The resulting analyses can produce very different results when examined with, for example, domain experts and beginners.

Many questions that needed answers, such as “how did the reward recipients feel”, went unanswered as the study did not have enough players who chose to receive a reward for analysis. We did not investigate whether or not the reward types differ between the versions (recreational and educational versions appear to have different types of rewards [9]). It is possible that more variety and more significance in the rewards may produce different results, as [10] and [11] suggest.

4.3 Comparison with previous research

Despite badges, points and achievements being "loved by most students" [8] and being the most valued gamification elements [8] the results of this study show that when compared to educational rewards the IMI survey shows no significant difference. The answers from all the IMI survey questions used (Tables 1, 2 and 3) show no significant differences when comparing the two different versions of the game. This finding contradicts [18] and supports the view that intrinsic motivation is not affected by points and badges, a finding shared with [19, 20].

Educational rewards show no changes in intrinsic motivation when compared to pointification rewards. This finding is the opposite of what [28] and [29]. Similarly, the use of difficulties or which version the student got to play had no impact on how far the student would progress in the game, a finding opposite of [14] and [27]. One possible explanation for these findings is that the players were simply "trying the game out" for a bit on their free time and had no further motivation past their curiosity. It is possible that these results came about because the study is not a longitudinal study or because the context of the bird quiz game is voluntary. In other contexts, such as a part of schoolwork, the results may differ. The 'pointification' rewards seem to be on the same level in the different IMI survey responses, a finding different from [34] where educational rewards were found to be less impactful. In addition, the results of this study differ with the findings from [35] where educational rewards were motivational to the students. However, in this case it needs to be said that they used cards for a much younger audience so the difference may be in the context.

The biggest differences that were found in this study were between the group of players that used difficulties (UD) and those who did not (DUD). The Table 11 shows that the UD group scored harder in statements that show that they tried harder on the activity, it was more important to do well at the task and they felt that they put less energy into this activity. In addition, the same group had statistically significantly different average total effort/importance scores, 23.5 for the UD and 19.1 for the DUD group. These findings can be interpreted to mean that if the players are given opportunities to increase the difficulty of their gameplay, the players feel that the activity required more effort and was more important to do well in. It is possible that the autonomy provided in the offering of increasing difficulty created additional intrinsic motivation and meaning in game play, supporting the view by [12] where it is stated that more research focus should be put into core game mechanics.

4.4 Future research

More practical research as suggested by [5] and the expansion of the gamification research with alternatives to 'pointification' as suggested by [32, 33]. This study can be replicated in different contexts. The ‘pointification’ rewards can be replicated nearly as is (points and badges), but the educational rewards should be different depending on the context. For example, if applied to learning the names of musical instruments, the educational reward of “additional information

about birds" can be replaced with "additional information about drums" which includes the history of drums, how they produce sound et cetera. Different quantities of rewards can also be tested to further explore the findings of [12]. In this study both game versions had the same number of different rewards available (unlike recreational and edu games found in [9]). This leaves open questions such as "does the number of rewards available have effect on intrinsic motivation?". Achieving motivated engagement is not easy [15] and needs further exploration in practical research. We recommend that the body of literature in this research field should be expanded with studies that measure outcomes with standardized surveys, similar to the recommendation of [13].

Theories other than the popular self-determination and flow theories [22] could be explored to find out how to harness the full potential of intrinsic motivation as described by [17]. More research that explores self-selected difficulty and its effects on motivation is necessary as this study supports the view that allowing people to select difficulties does increase the motivation of players for the players who chose to [16].

5. Conclusions

In this study we examined the impact of different rewards (educational and 'pointification') and optional use of in-game difficulties (used or did not use in game difficulties) to examine whether or not the players intrinsic motivation inventory scores would reveal changes in the player experience. The results show that the educational and pointification versions of the game had no statistically significant difference in the player experience as measured by the IMI survey. Based on these findings, we recommend that future game designs experiment with rewards more closely tied to the subject matter or context of the game, as generic rewards may not sufficiently engage players.

However, players who engaged with optional difficulties reported significantly higher effort/importance scores. We recommend integrating customizable difficulty settings, as they can enhance players' intrinsic motivation. This creates promise in the future research of optional difficulties in the field of gamification. The effects of optional difficulty options can be further explored with different surveys and analyses. In addition, the findings show that the reward type nor engagement with optional difficulty settings had effect on how far the players progressed in-game. This suggests that reward types available or the use of difficulty settings does not affect how far the players will progress in-game. We suggest that the game elements that affect in-game progression are explored further in future research.

This study is relevant in the field of gamification, serious-, and educational game design. The significance of this study can be seen in the novel ways to test out different types of features (such as 'pointification' and educational rewards or the use of difficulty systems) and their effects on the intrinsic motivation inventory survey. These findings can be used to build further experimentation and to guide the efforts of future researchers.

In conclusion, this study contributes to the growing body of knowledge more information about the intrinsic motivation of the players and how different rewards and the addition of optional difficulties can change the player experience. The results highlight the need for additional studies for alternatives for the points and badges in addition to the need for more rigorous empirical study for core gameplay mechanics and their effects on the intrinsic motivation and the player experience in general.

Acknowledgments

The bird quiz game in this study was created using the Unity Game Engine. Some of the visible assets were purchased via the Unity Asset Store from the publisher "Layer Lab" (https://assetstore.unity.com/publishers/5232). Permission for the usage of assets for the research was asked from Unity directly.

Special thanks to Kimmo Tyni for the bird images used in the game.

Conflicts of interest

The authors report no possible conflicts of interest.

References

[1] S. Deterding, R. Khaled, L. Nacke, & D. Dixon, "Gamification: Toward a definition," in Proceedings of CHI 2011 Workshop Gamification: Using Game Design Elements in Non-Game Contexts, 2011, pp. 6–9, doi: 10.1145/1979742.1979575.
[2] J. Dah, N. Hussin, M. K. Zaini, L. I. Helda, D. S. Ametefe, A. A. Aliu, W. Suqi, & A. Caliskan, " Gamification Equilibrium: The Fulcrum for Balanced Intrinsic Motivation and Extrinsic Rewards in Learning Systems," IJSG, vol. 10, no. 3, pp. 83–116, 2023, doi: 10.17083/ijsg.v10i3.633.
[3] J. Koivisto & J. Hamari, "The rise of motivational information systems: A review of gamification research," International Journal of Information Management, vol. 45, pp. 191–210, 2019, doi: 10.1016/j.ijinfomgt.2018.10.013.
[4] A.-S. Hellberg & J. Moll, "A point with pointsification? Clarifying and separating pointsification from gamification in education," Frontiers in Education, vol. 8, p. 1212994, 2023, doi: 10.3389/feduc.2023.1212994.
[5] J. Tyni, A. Tarkiainen, S. López-Pernas, M. Saqr, J. Kahila, R. Bednarik, & M. Tedre, "Games and rewards: A scientometric study of rewards in educational and serious games," IEEE Access, vol. 10, pp. 31578–31585, 2022, doi: 10.1109/ACCESS.2022.3160230.
[6] S. Hallifax, M. Altmeyer, K. Kölln, M. Rauschenberger, & L. E. Nacke, "From points to progression: A scoping review of game elements in gamification research with a content analysis of 280 research papers," Proceedings of the ACM on Human-Computer Interaction, vol. 7, no. CHI PLAY, pp. 748–768, 2023, doi: doi.org/10.1145/3611048.
[7] J. Zeng, S. Parks, & J. Shang, "To learn scientifically, effectively, and enjoyably: A review of educational games," Human Behavior and Emerging Technologies, vol. 2, no. 2, pp. 186–195, 2020, doi: 10.1002/hbe2.188.
[8] M. Garcia-Iruela & R. Hijón-Neira, "What perception do students have about the gamification elements?," IEEE Access, vol. 8, pp. 134386–134392, 2020, doi: 10.1109/ACCESS.2020.3011222.
[9] J. Tyni, A. Turunen, J. Kahila, R. Bednarik, & M. Tedre, "Reward types in popular recreational and educational mobile games," IEEE Access, vol. 11, pp. 1166–1174, 2022, doi: 10.1109/ACCESS.2022.3231936.
[10] C. Phillips, D. Johnson, M. Klarkowski, M. J. White, & L. Hides, "The impact of rewards and trait reward responsiveness on player motivation," in Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play, pp. 393–404, 2018, doi: 10.1145/3242671.3242713.
[11] D. Johnson, M. Klarkowski, K. Vella, C. Phillips, M. McEwan, & C. N. Watling, "Greater rewards in videogames lead to more presence, enjoyment, and effort," Computers in Human Behavior, vol. 87, pp. 66–74, 2018, doi: 10.1016/j.chb.2018.05.025.
[12] B. McKernan, R. M. Martey, J. Stromer-Galley, K. Kenski, B. A. Clegg, J. E. Folkestad, M. G. Rhodes, A. Shaw, E. T. Saulnier, & T. Strzalkowski, "We don’t need no stinkin’ badges: The impact of

reward features and feeling rewarded in educational games," Computers in Human Behavior, vol. 45, pp. 299–306, 2015, doi: 10.1016/j.chb.2014.12.028.
[13] J. Hamari, "Do badges increase user activity? A field experiment on the effects of gamification," Computers in Human Behavior, vol. 71, pp. 469–478, 2017, doi: 10.1016/j.chb.2015.03.036.
[14] S. Fodor & B. Balázs, "An empirical study on key factors affecting user engagement in a gamified team building environment," IJSG, vol. 7, no. 3, pp. 81–95, Sep. 2020, doi: 10.17083/ijsg.v7i3.355.
[15] T. H. Laine & R. S. N. Lindberg, "Designing engaging games for education: A systematic literature review on game motivators and design principles," IEEE Transactions on Learning Technologies, vol. 13, no. 4, pp. 804–821, 2020, doi: 10.1109/TLT.2020.3018503.
[16] J. D. Lomas, K. Koedinger, N. Patel, S. Shodhan, N. Poonwala, & J. L. Forlizzi, "Is difficulty overrated? The effects of choice, novelty, and suspense on intrinsic motivation in educational games," in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, pp. 1028–1039, 2017, doi: 10.1145/3025453.3025638.
[17] R. M. Ryan & E. L. Deci, "Intrinsic and extrinsic motivations: Classic definitions and new directions," Contemporary Educational Psychology, vol. 25, no. 1, pp. 54–67, 2000, doi: 10.1006/ceps.1999.1020.
[18] J. Xu, A. Lio, H. Dhaliwal, S. Andrei, S. Balakrishnan, U. Nagani, & S. Samadder, "Psychological interventions of virtual gamification within academic intrinsic motivation: A systematic review," Journal of Affective Disorders, vol. 293, pp. 444–465, 2021, doi: 10.1016/j.jad.2021.06.070.
[19] D. Dicheva, R. Caldwell, & B. Guy, "Do badges increase student engagement and motivation?," in Proceedings of the 21st Annual Conference on Information Technology Education, pp. 81–86, 2020, doi: 10.1145/3368308.3415393.
[20] E. D. Mekler, F. Brühlmann, K. Opwis, & A. N. Tuch, "Do points, levels, and leaderboards harm intrinsic motivation? An empirical analysis of common gamification elements," in Proceedings of the First International Conference on Gameful Design, Research, and Applications, pp. 66–73, 2013, doi: 10.1145/2583008.2583017.
[21] E. D. Mekler, J. A. Bopp, A. N. Tuch, & K. Opwis, "A systematic review of quantitative studies on the enjoyment of digital entertainment games," in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 927–936, 2014, doi: 10.1145/2556288.2557078.
[22] J. Krath, L. Schürmann, & H. F. O. Von Korflesch, "Revealing the theoretical basis of gamification: A systematic review and analysis of theory in research on gamification, serious games, and game-based learning," Computers in Human Behavior, vol. 125, p. 106963, 2021, doi: 10.1016/j.chb.2021.106963.
[23] R. M. Ryan, V. Mims, & R. Koestner, "Relation of reward contingency and interpersonal context to intrinsic motivation: A review and test using cognitive evaluation theory," Journal of Personality and Social Psychology, vol. 45, no. 4, pp. 736–750, 1983, doi: 10.1037/0022-3514.45.4.736.
[24] Intrinsic Motivation Inventory (IMI)," Center for Self-Determination Theory. Accessed on: 11.05.2024. [Online]. Available: https://selfdeterminationtheory.org/intrinsic-motivation-inventory/
[25] A. Drachen, P. Mirza-Babaei, & L. E. Nacke (Eds.), Games User Research, Oxford University Press, 2018, doi: 10.1093/oso/9780198794844.001.0001.
[26] A. J. Nieuwhof-Leppink, T. P. V. M. de Jong, E. M. Van de Putte, & R. Schappin, "Does a serious game increase intrinsic motivation in children receiving urotherapy?," Journal of Pediatric Urology, vol. 15, no. 1, pp. 36-e1, 2019, doi: 10.1016/j.jpurol.2018.09.003.
[27] I. Gurjanow, M. Oliveira, J. Zender, P. A. Santos, & M. Ludwig, "Mathematics trails: Shallow and deep gamification," International Journal of Serious Games, vol. 6, no. 3, pp. 65–79, 2019, doi: 10.17083/ijsg.v6i3.306.
[28] T. Cassells & D. O. Brain, "The effect of gamification on intrinsic motivation for prioritisation," in 2018 IEEE Games, Entertainment, Media Conference (GEM), IEEE, pp. 1–11, 2018, doi: 10.1109/GEM.2018.8516517.
[29] A. Peter, C. Salimun, & E. A. A. Seman, "The effect of individual gamification elements in intrinsic motivation and performance," Asian Journal of Research in Education and Social Sciences, vol. 1, no. 1, pp. 48–61, 2019, doi: 10.1088/1742-6596/1358/1/012058.
[30] R. Klemke, A. Antonaci, & B. Limbu, "Designing and implementing gamification: GaDeP, Gamifire, and applied case studies," IJSG, vol. 7, no. 3, pp. 97–129, 2020, doi: 10.17083/ijsg.v7i3.357.

[31] N. Xi & J. Hamari, "Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction," International Journal of Information Management, vol. 46, pp. 210–221, 2019, doi: 10.1016/j.ijinfomgt.2018.12.002.
[32] T. Barik, E. Murphy-Hill, & T. Zimmermann, "A perspective on blending programming environments and games: Beyond points, badges, and leaderboards," in 2016 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC), pp. 134–142, 2016, doi: 10.1109/VLHCC.2016.7739676.
[33] J. Dah, N. Hussin, M. K. Zaini, L. I. Helda, D. S. Ametefe, & A. A. Aliu, "Gamification is not working: Why?," Games and Culture, pp. 15554120241228125, 2024, doi: 10.1177/15554120241228125.
[34] L. Whittaker, R. Russell-Bennett, & R. Mulcahy, "Reward-based or meaningful gaming? A field study on game mechanics and serious games for sustainability," Psychology & Marketing, vol. 38, no. 6, pp. 981–1000, 2021, doi: 10.1002/mar.21476.
[35] P. Chen, R. Kuo, M. Chang, & J.-S. Heh, "The effectiveness of using in-game cards as reward," Research and Practice in Technology Enhanced Learning, vol. 12, no. 1, pp. 1–23, 2017, doi: 10.1186/s41039-017-0054-8.
[36] D. Dicheva, C. Dichev, G. Agre, and G. Angelova, "Gamification in education: A systematic mapping study," *Journal of Educational Technology & Society*, vol. 18, no. 3, pp. 75-88, 2015. [Online]. Available: https://www.jstor.org/stable/jeductechsoci.18.3.75