Computer Applications in Engineering Education

WILEY

RESEARCH ARTICLE

Studying the Effectiveness of Games as an Extracurricular Activity in a Higher Education Programming Course

Silvia Arias-Herguedas | Alma Pisabarro-Marron | Carlos Vivaracho-Pascual | Alejandro Ortega-Arranz | Luis Ignacio Jiménez

Department of Computer Science, University of Valladolid, Valladolid, Spain

Correspondence: Carlos Vivaracho-Pascual (cevp@infor.uva.es)

Received: 6 May 2024 | **Revised:** 7 August 2024 | **Accepted:** 7 January 2025

Funding: This work was supported by Ministerio de Ciencia, Innovación y Universidades.

Keywords: computer science | extracurricular activity | fundamentals of programming | games | higher education

ABSTRACT

Higher Education programming courses usually present high levels of student failure and drop-out rates. Given this context, the use of educational video games is proposed as a strategy to increase the students' motivation and engagement, thus helping diminish such rates. However, there is a lack of empirical studies examining such effects, especially when they are proposed outside the formal curriculum (i.e., extracurricular) and when the students are enrolled in different Higher Education bachelors. This paper presents a cross-sectional study following a between-subjects design with 315 students (168 assigned to the experimental condition, 147 to the control condition) enrolled in one of the following bachelors: B.S. in Computer Science (where programming is a core subject), and B.S. in Statistics (where it is not). The study spanned two consecutive academic years. The outcomes were evaluated through a pre-/post-test schema and comparison of final course results to measure the effect on learning (objective assessment) and a survey to get the students' perceptions (subjective assessment). In addition, the level of participation was analyzed and compared between bachelors, considering the optional nature of the activity. Results show statistically significant differences in learning outcomes between the students in the experimental condition and those in the control group, without clear differences between Bachelors (the results are positive for both). In the subjective assessment and participation, the results are also positive, but, in this case, statistically significant differences between bachelors have been observed. These positive outcomes suggest its potential applicability to other Higher Education and Engineering courses.

1 | Introduction

Nowadays, video games are one of the most popular forms of entertainment among the worldwide population [1]. Video games, and in the broader sense, games, are able to stimulate the feelings of the players (e.g., fun, enthusiasm, competitiveness), encouraging them to achieve the expected game goals [2]. This feature makes games powerful tools that can be used to promote goal achievement in non-recreational contexts, for example, in education [3]. The term serious game was coined to refer to those games whose main purpose is different from mere entertainment [4]; more specifically, those serious games whose main goal is pedagogical are termed educational games [5].

Educational games have been proven to increase the students' levels of motivation and learning outcomes at different educational levels [6, 7]. Previous researchers frame these benefits within existing psychological theories, such as the *Flow Theory* [8, 9] and the *Self-Determination Theory* [10, 11]. These theories propose a proper balance between the personal capabilities of the students and the established game challenges (i.e., the *Flow Theory*); as well as the satisfaction of three basic human needs (i.e., autonomy, competence, and relatedness) to keep students intrinsically motivated within the game (i.e., the *Self-Determination Theory*).

Consequently, educational games following these principles are expected to enhance student motivation and improve the learning outcomes. This holds particularly true for Higher Education programming courses, which often exhibit high levels of student failure and drop-out rates [12, 13]. This challenge is exacerbated when dealing with first-year university students who struggle with abstract programming concepts for the first time, such as variables, loops, stacks, and linked lists [13, 14]. Nevertheless, the rigid curriculum of these courses poses difficulties to integrate educational games as a formal part of the curriculum; this is even more difficult if the use of games is to be encompassed within the entire curriculum taught throughout the course, as in the activity proposed here (Section 3). Hence, a possible solution for incorporating these games into such courses is to recommend their use beyond regular class hours (i.e., as extracurricular activities).

In the last decade, there has been an increasing number of studies attempting to understand the (positive and negative) effects of educational games within programming courses at different educational levels [15–18]. In summary, the related literature (Section 2) emphasizes the positive value of educational games within the formal context of programming courses. Nevertheless, many of these studies address the impact of educational games on students' learning outcomes only from a self-perceived perspective and without comparing the effects of playing and not playing. The need of "Further empirical research to better understand the potential impact and effective education implementation of video games in computer science" was already shown in [19], although it is still an open question, as is confirmed in more recent reviews [16]. In addition, these previous studies usually propose the use of games as part of the formal curriculum of the course, thus forcing students to play games even if they are not engaging for them. So, the impact of the self-government use of games by students outside curricular hours has not been sufficiently studied. To address these previous concerns, this study provides empirical evidence regarding the influence of extracurricular educational games in a first-year Higher Education programming course. Accordingly, this study aims to answer the following research questions:

RQ1. *To what extent does the use of an extracurricular game-based activity foster the student learning outcomes in a Higher Education programming course?*

RQ2. *What are the differences in the learning outcomes between students pursuing a Bachelor in Science (BS) in Statistics and those pursuing a BS in Computer Science?*

Additionally, the motivation outcome is also approached in order to study whether the activity increased it (**RQ3**), as well as the students' perceptions towards the activity, that is, if they enjoyed it or not (**RQ4**). So, the study is focused on assessing the impact of the activity on the emotional (user experience), behavioral (motivation), and cognitive (learning) competences of the students.

To address these questions, a between-subjects study [20] was conducted during two consecutive academic years, considering experimental and control conditions, composed of students proposed to participate in the activity and those who were not, respectively. To report this quantitative research, the paper structure is based on the JARS-Quant (Journal Article Reporting Standards for Quantitative Research) template [21].

As pointed out, both our experience and the literature [12–14] allow us to state that programming is a complex and difficult subject for novice students. So, we consider it interesting to propose alternatives that foster students' learning while their experience is also improved, as in the present study.

The rest of the manuscript is as follows. Section 2 describes and compares similar research studies dealing with the effect of educational games on programming students. Section 3 introduces the extracurricular game-based activity proposed. Section 4 presents the research design of the study, including the participants and data sources. Section 5 reports the results of the study. Section 6 discusses the findings. Sections 7 and 8 describe the threats to validity and limitations of the study, respectively. Finally, Section 9 outlines the main conclusions of this work and the proposed future extensions.

2 | Related Works

Proficiency in programming stands as a fundamental pillar in contemporary society [12]. Its significance extends to bolstering technological progress, fostering innovation, and contributing to economic expansion [22]. Expertize in programming is widely regarded as indispensable for securing employment across diverse industrial sectors [23]. Furthermore, it cultivates adept problem-solving capabilities and nurtures creativity, enabling individuals to streamline tasks and significantly enhance operational efficiency [15]. However, in many cases, the students perceive programming as a complex topic [12, 13, 24], and it is difficult to engage them [25].

Given this context, the use of educational games for programming has experienced a notable increase in recent years at all educational levels [16, 17, 26, 27]. Although the majority of the approaches developed their own games or proposals, publicly available ones can be found in commercial stores or in open can be found [1, 28, 29]. In the last few years, Virtual Reality has been introduced with the aim of producing a more immersive sensations [30–32], but with no clear advantages.

Many works in the related literature do not show the specific programming concepts involved in the games proposed and/or tested; they "speak" of programming concepts or computational/logical thinking in general. From the works that explicitly show this, two different approaches can be found:

- Games focused on a specific concept, such as [25], where Minecraft is used in a 1-h session to practice the sequence control structures, [33] where the game approaches the concept of iteration, [34] where a game is designed aimed at understanding the stack concept, or in [35] where the difficult concept for novices of recursion is approached.

- Games that include several programming concepts. In [1], a list of public games can be seen with the programming concepts involved in each one. Other examples can be seen in [36] or [1]: in the first, the same game involves several learning outcomes, while in the second, the game includes several *minigames*, each one focused on different programming concepts.

Our approach is that in the first item, but integrating different and independent *minigames* in the same activity (Section 3), this being developed along the academic year and covering the main concepts of the subject. Each *minigame* focuses on a particular programming concept, so each one is worked separately and deeply. This is an original approach, to the best of our knowledge, having found the following three most relevant similar approaches.

The most similar is that shown in the works of Zhao et al. [13, 23], developing and testing three different games in different university academic contexts, focused on loops, functions, and abstract concepts of structure in C Programming, respectively. In [13], a qualitative evaluation based on a survey is shown, while in [23], a quantitative evaluation of the students' learning outcomes is performed based on ad-hoc pre- and post-tests. This last evaluation is carried out for each game independently, showing a statistically significant improvement in the corresponding programming competence. The number of students in each game evaluation was 54, 66, and 67, respectively, all in a unique experimental set, that is, a control group was not used. This is a difference with regard to the study shown here. Another difference is that, as pointed out, here the games are treated as “a whole,” evaluating their influence in the learning outcomes by means of curricular course tests. It can be considered that this work complements and reinforces the results of Zhao et al.

Another similar approach can be seen in [1], where a set of *minigames* are integrated in the proposed *Code-Venture* game. Only a qualitative assessment was performed, collecting the students' opinions about the game and their perception of programming before and after playing. The results showed a good valuation of the game and an increase in the positive feelings about programming. The sample size was only 35 participants between 16 and 21 years old.

The third work is [36], where several games related to programming are shown. Each game is completely different and is not related to the others with regard to the learning outcomes. These games were developed for the final-year students at the University of the West of Scotland as a project to develop games for courses they had undertaken in earlier years. Evaluation is not performed; only a description of each game is shown.

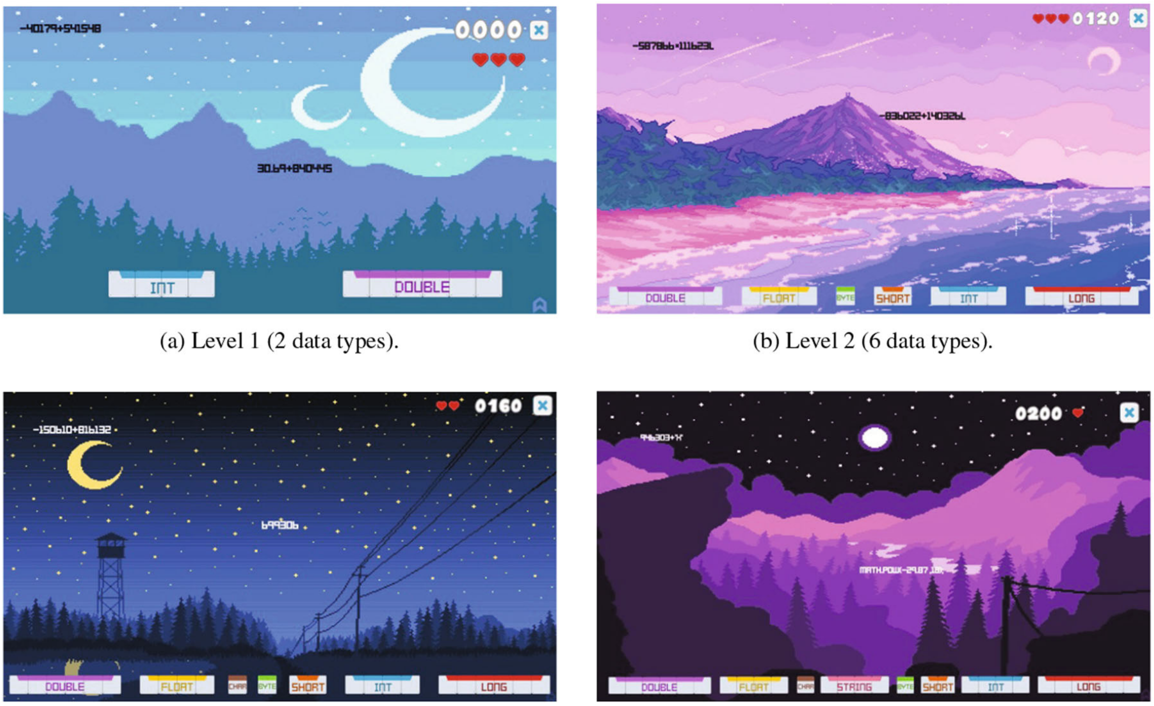
Another important characteristic of our proposal is its extracurricular character, having not found in the related literature a similar approach in this sense. Focusing on “games AND extracurricular AND education,” in general, works related to physical activity or exergames (e.g., [37, 38]) or with non-digital games (e.g., [39, 40]) can be found; although the last two references are related with university education, the majority of the proposals are developed for previous educative stages, including those related with programming [41, 42]. Once more, the most relevant similar studies using, in this case, games as an extracurricular activity in University are described in the following.

In [43], a videogame about the Complete Blood Count analysis for healthcare students is developed and tested. The goal of the game is to complement the classic lecture-based approach and improve their knowledge and skills in hematology. So, a single extracurricular session of 1.5 h was set up for a group of 153 volunteers, 86 (of 324 students) of Medicine and 67 (of 115 students) of Pharmacy degrees. Pre-game and post-game tests were performed, showing an increase in the grades in the post-game test. Satisfaction surveys were also recorded immediately after the activity, with a participation of 143 students; 86% of them answered that they had strengthened their knowledge and 80% had had fun. In addition, the final exam grades of the students who played and did not play were compared, showing a better performance of the students in the first set, this being statistically significant for the students of Medicine, but not for the Pharmacy ones; no pre-test as baseline was used for this comparison. Another important difference with the present study is that the experimental sets were not selected randomly.

A set of role-playing didactic games (no videogames) are used in ref. [44], implemented in extracurricular activities to form the legal competence of the students with a bachelor's degree in the field of psychological and pedagogical education. To assess the proposal, the students were divided into an experimental set ($N = 94$) (played the games) and a control set ($N = 90$) (did not play the games). The students were classified according to their competence in three levels (high, average, and low) using the grades in pre-activity and post-activity tests. The results showed a better and statistically significant performance (the number of students in the high and average competence levels increased) of the students in the experimental set in the majority of formed legal competencies. The activity proposed can be understood as more “extra” than “outside” the curriculum, as our proposal is.

Although it is not directly related to our study, in ref. [45] an interesting initiative for Computer Science students, RadGrad, is presented. It is an online application combining features of social networks, degree planners and serious games. Its goal is that the students go beyond curricular activities to broaden their knowledge in the field and promote participation in extracurricular activities, thus improving engagement, diversity, and retention.

As has been shown, there is a lack of empirical studies dealing with the effects of extracurricular game-based activities (optional, outside the regular teaching hours and with no impact on the subject grades) in programming in particular, and the studies concerning the effects on university education, in general, are scarce and limited. In the same sense, not many works can be found that propose and assess game-based activities that encompass the entire curriculum taught throughout the course. Additionally, none of the previous studies have explored the existence of potential differences according to participants' different contexts in the same subject, that is, in the same educative context; [13] approaches a similar comparative study, but the students belong to different universities. To this end, this work aims to approach these issues.



(a) Level 1 (2 data types). (b) Level 2 (6 data types).

(c) Level 3 (7 data types). (d) Level 4 (8 data types). Besides, the platforms are rising.

FIGURE 1 | Screenshots of the different levels implemented, with incremental difficulty, in the *Data Drop minigame*. Players move left and right the falling quantities to make them match with their type.

3 | Game-Based Activity Description

The game-based activity consists of a series of educational video games illustrating different programming concepts either in a literal or metaphorical manner. The games are developed to require both a short amount of time to successfully complete them and a simple interaction (just a mouse click); for this, they are called *minigames*. The *minigames* were designed to engage the students, encouraging them to play, following an *Emotional design* [12].

Each *minigame* is specifically crafted to emphasize a basic programming concept, either through literal or metaphorical representation. The *minigames* feature different game mechanics (e.g., points, timers). The current roster of *minigames*, along with their associated programming concepts, is:

1. *Data Drop*. To identify and differentiate the different Java data types.
2. *Mars Miners*. To master flow control in structured programs.
3. *Recursive Party*. To understand the concept of recursion.
4. *Pointed*. To learn dynamic data structures (oriented to linked lists).
5. *Stacked*. To learn dynamic data structures (oriented to stacks and queues).

The games cover the programming basics in imperative programming pointed out in [12].

Each *minigame* becomes accessible to students only at the discretion of the teacher, approximately every 2 weeks (Appendix A). This approach ensures that the *minigames* are made available as the subject matter advances. Consequently, students engage with the *minigames* once they have already grasped the relevant concepts.

Most *minigames* mechanics are common and based on progression, such as scoring, life counting or stratification into varying levels of difficulty. The concept of *flow channel* [46] is established by the gradual introduction of increasing levels of difficulty, which are interconnected with the gameplay itself (to be dynamic to the player's skill level [12]) and also with the underlying programming principles implemented in the game.

All levels have an independent score. Consequently, players have the opportunity to accumulate points even if they do not complete the *minigame*. Following the *flow channel* concept, an upper level cannot be accessed until the successful completion of all preceding levels. This gradual unlocking of levels allows students to grasp the fundamental concepts before moving to more difficult ones. For instance, consider the *minigame Data Drop*, which elucidates primitive data types in Java. Each level of this *minigame* systematically introduces new data types, starting with basic numeric types, such as integer or double, and finishing with more complex ones, such as Strings. Figure 1 depicts screenshots of various levels of this *minigame*.

The participation in *minigames* was optional and to be played outside the regular teaching time, owing to the aforementioned constraints that exclude its inclusion within the formal course curriculum. This approach entails a twofold advantage for this research. Firstly, the students are not forced to participate, thus helping us discriminate those students that actually like the

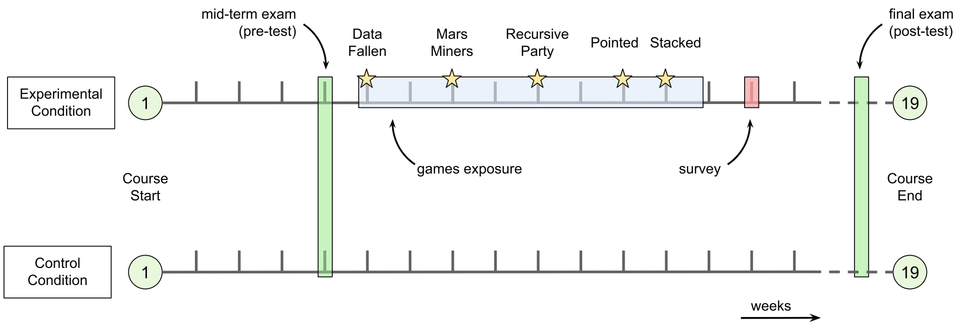


FIGURE 2 | Graphical visualization of the research design of the study. The first line involves students who participated in the activity, and the second line to those who did not.

minigames; and secondly, the students' activity performance does not affect their score in the course, thus ensuring that the students' behavior is not affected by this external motivator.

The activity's plot thread is the reconstruction of a lost constellation. Therefore, the students need to get back the stars lost from the constellation. Stars are earned by completing the *minigames*. To complete them (i.e., mark them as passed), students must get a minimum score in each game, thus ensuring that the basic concepts of the *minigames* have been learned. Once the *minigame* is passed, students can play as many times as they wish to improve the *minigame's* score and time. Nevertheless, stars are only issued the first time a *minigame* is completed disregarding the number of times the score is improved.

Individual and group leaderboards are configured to rank students and groups (groups were randomly formed based on the university cohorts created for teaching this course) according to the number of stars and points earned (in case of equal number of stars and points between two students or groups, the time to complete the *minigame* is used to rank students). Group leaderboards show the sum of stars and points earned by all members of the same group for every *minigame*, thus promoting their individual accountability, participation, and cooperation. Thus, three different challenges were posed to the participants throughout the activity: (1) individual challenge to reach the first positions of each *minigame*, (2) individual challenge to reach the first positions by adding together up all the *minigames*, and (3) a group challenge to reach the first positions adding up the stars and scores from all group members. The winners of each of these competitions receive a prize that, as pointed out, is not related with the subject. All the games and leaderboards were hosted in the *GamiSpace* platform. A demo of the platform and the games is available at: http://demogamispace.infor.uva.es/.

4 | Methodology

This study follows a *Cross-Sectional* **design** [47, 48], based on a *between-subjects* approach [20] with experimental and control conditions. Both qualitative and quantitative data were collected.

The qualitative part of the study focused on the students' *reaction* (Kirkpatrick's model level 1) [49]. Data were gathered by means of a survey administered to the experimental condition as commonly done in this type of studies [50].

The quantitative part aims to objectively analyze the influence of the game-based activity in the learning process or competence acquisition in the subject (Kirkpatrick's model level 2) [49]. To this end, the study follows a pre-/post-test schema [50, 51], where the pre-test (carried out before the activity) is used as a baseline with which to compare the later evolution of the students' learning through the post-test. For both pre- and post-tests, curricular course tests are used. In addition, the final results from the course of the experimental and control sets are also compared.

A visual summary of the research design is shown in Figure 2, which also includes the *minigames* and evaluation activities schedule (Appendix A).

The **participants** of the study were 377 students enrolled either in a BS in Computer Science ($N = 302$) or a BS in Statistics ($N = 75$) of the University of Valladolid (Spain), all of them taking the same course on computer programming. The study was performed during two consecutive academic years, as in [34, 52]. Computer programming is a first-year course in both bachelors, so participants' ages mostly range between 18 and 20. Additionally, all participants were from the same country, and their prior knowledge was similar, since the great majority came from the same previous educational stage (except for a small percentage of students who are re-taking the course). Before the study, all the participants were informed of its purpose, and it was guaranteed that their participation or withdrawal would incur neither reward nor punishment.

The participants were **sampled** as follows. They were divided into two experimental (participating in the game-based activity in the BS in Computer Science and in the BS in Statistics), and two control conditions (did not participate in the game-based activity for both bachelors) each academic year. Participants were assigned to the aforementioned conditions following a simple random process based on the cohorts created to teach this course. The experimental condition involved an initial total number of 206 students (163 BS in Computer Science, 43 BS in Statistics); and the control condition consisted initially of 171 students (139 BS in Computer Science, 32 on Statistics). These numbers satisfy the minimum number of participants per condition suggested to assess the effectiveness of game-based learning strategies [51]. As the activity is voluntary, some of the initial components of the experimental condition did not participate in it, that is, they did not play any game. Therefore, the activity did not impact the behavior, cognition, or emotion of

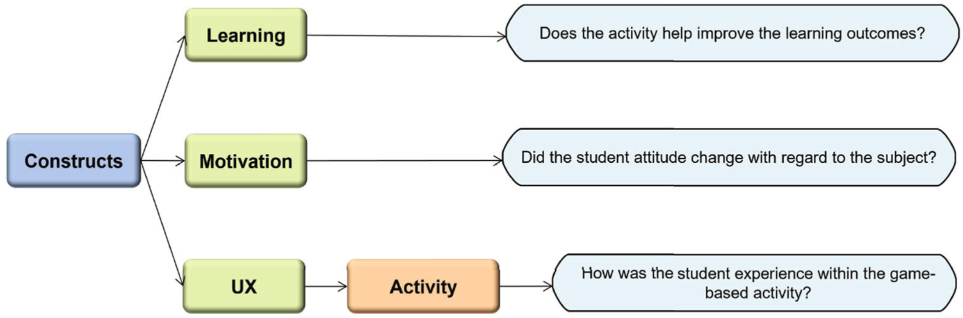


FIGURE 3 | Structure of the model used to evaluate the activity.

these students, so they were eliminated from the study. In Section 5, a clearer description of the participant flow is shown.

To achieve the study objectives, what to **measure** must first be defined. This was derived from the goals, gathered in the research questions (Section 1). This definition was made based on the well-defined Model for the Evaluation of Educational Games (MEEGA) [53, 54]. Following this model, the measures were defined by means of the hierarchical decomposition shown in Figure 3. The *motivation* construct or sub-component was adapted, since the original model measures the student attitude with regard to the game, and here we want to measure it with regard to the subject. The construct *User Experience (UX)* measures the interaction of the students with the activity; from the different dimensions proposed in the model, we are only interested here in the emotional response evoked in the player (*amusement dimension* in the model). The *learning* construct is the most important in this study, since it measures an essential goal in educational games, which is the effect on the students' competence; from the dimensions proposed, the study focused on the so called *short term learning*, which is related to the learning goals of the subject (here, fundamentals of programming).

To evaluate the activity outcomes, a **hypotheses assessment** is followed. The independent variable of the study was "participating in the activity." The dependent ones, aligned with the constructs decomposition shown, were learning, motivation, and the student's experience. Following the Cross-Sectional design, no other independent variables were controlled. However, their influence was minimized, as described in Section 7.2. So, the following research hypotheses were stated:

- **Primary hypothesis 1 (PH1):** The activity boosts student learning.
- **Primary hypothesis 2 (PH2):** The activity increases student motivation.
- **Primary hypothesis 3 (PH3):** The students liked the activity.

Although all the outcomes related with the earlier hypotheses are important, since they have a positive influence over the students, the established order underscores significance within the study due to its impact on the students.

4.1 | Data Collection

Appendix A sets out the course, *minigames,* and study activities schedule to better understand the timing of the data collection and its relation with that of the course activities and *minigames* availability.

4.1.1 | Objective Assessment

The influence of the game-based activity in the learning process was evaluated using a pre-test and a post-test [51] schema and comparing the final results of the course.

For the pre- and post-tests, as pointed out, curricular course tests were used (available in https://greidi.infor.uva.es/CAEE_GamAsExtAct_AddMat/Pre-Post_Tests.zip). The pre-test (PrT) consisted of a test administered to the students in the fifth week of the semester and before the start of the activity. In this test, students are examined on the content taught up to that point in the subject (Appendix A). In this study, this test aimed to get a reference with which compare the evolution of the students. As post-test, the final course test was used (Appendix A). This includes all the programming concepts, and is done after having finished both the course and the activity. The final grade of the subject depends mainly on the mark in this test.

The final course test is also used to measure the final competence of the students in the subject, in order to compare it using the control and experimental sets.

4.1.2 | Survey

The design of the survey followed the suggestion in [54, 55] and was based on previous similar ones of the authors [56], which were matched with well-defined scales [54, 57, 58].

When gathering student opinion, besides that related with the study in this work, there also existed interest in collecting information related to the games platform and each minigame in order to continuously improve this part of the activity (both surveys are available in https://greidi.infor.uva.es/CAEE_GamAsExtAct_AddMat/Surveys.zip). In addition, it was important not to overload the students (they have more surveys related with other subjects, besides others related with institutional evaluations) and to minimize the intervention over the normal development of the subject. So, it was decided to include all in a single survey.

In order to keep this survey short, it included only three items (Figure 7) related with this study, which were based on a previously designed questionnaire [56]; each one is directly related with the different measures presented in Figure 3. Items can be answered in a 5-point Likert-like scale ranging from a very negative opinion to a very positive one. The scale was masked so that the answers were not numbered.

The survey was anonymous and was fitted on a single sheet of paper. As with the responses, the items were not numbered. It was taken in the last week of the course (Appendix A), prior to the final exam and once the activity had finished.

4.2 | Statistical Methods

4.2.1 | Objective Assessment

This analysis addresses the answer to RQ1, by assessing PH1, and to RQ2, objectively.

- **Pre-Post Test Analysis** As the pre-test is the reference or baseline, the post-test (Section 4.1) results were compared with this, calculating the difference between the marks in both: $Dif = \text{PostTest\_Grade} - \text{PreTest\_Grade}$. The relative value was also obtained: $RelDif = \frac{\text{Dif}}{\text{PreTest\_Mark}}$. The mean of the values for each student was calculated and compared to those of the experimental and control sets.

  The *Mann–Whitney–Wilcoxon test* (Mann–Whitney U-Test) was used to measure the statistical significance of this difference. This test was selected because the *Dif* and *RelDif* values follow non-parametric distributions, are ordinal variables, and the participants were randomly assigned to the different conditions (independent samples). The null hypothesis to evaluate is H_0: *the relation between marks in the pre- and post-tests have the same distribution in the experimental and control sets*.
- **Course Final Results Analysis** The mean of the grades in the final test were calculated for both the experimental and control sets. In addition, the failure and approval rates for both sets were also calculated. The grades range from 0 to 10, 0 being the worst and 10 being the best. A student passes if a minimum grade of 5 is achieved.

  In the same way as with the pre-/post-test analysis, the statistical significance of the differences was calculated by means of the *Mann–Whitney–Wilcoxon test*.
- **Data diagnostics** The previous analyses were done for all the students, as well as grouping them by BS (Computer Science and Statistics students) to study the effect of this "parameter." In addition, this analysis was performed per academic year and joining both for a wider study.

4.2.2 | Survey Analysis

This analysis addresses the answer to RQ1, RQ3, and RQ4, by assessing PH1 (but now based on the students' opinion), PH2 and PH3, respectively. The statistical study was supported by data visualization for a more complete and understandable description of the results.

Each activity outcome was posed by means of a primary hypothesis. These are validated estimating the probability of achieving positive results (values 4 or 5 in the responses) in each survey question, together to its confidence interval at 95%. An evaluation was performed to determine whether the probability of obtaining positive results was not achieved by chance, using a test for $p, H_0: p = 0.5$, that is, the events (here, the probability of positive results) happened randomly. The conclusion of this test is shown in the column "Significant?" (if H_0 is refused or not) of the results tables. In the case of *non significant* results, the *Test Power* [59] was calculated.

The comparative analysis of the responses of each BS (Computer Sciences vs. Statistics) was performed. Since these responses do not follow a normal distribution, are ordinal variables, and the samples are independent, the non-parametric *Mann–Whitney–Wilcoxon test* was also used to analyze the statistical significance of the differences, but now with H_0: *the survey responses of the Computer Science BS and Statistics BS students have the same distribution*.

5 | Results

Figure 4 shows, based on JARS-Quant schema [21], the **the participants flow**, that is, the total number of participants in each group at each stage of the study.

As for the values depicted in the green squares of Figure 4, related to the analysis part of the study, it is important to point out that the survey was only considered for the students in the experimental set, therefore this does not appear in the control set part. Although the survey was voluntary, it was answered by a high percentage: 75% (73 of 98) the first year and 91% (71 of 78) the second. The same occurs with the pre-/post-test study, where the students that dropped out of the subject were not considered (they had no grades), but this number is, in general, low both in the experimental and in the control set. Since the analysis was performed with all the students, as well as grouping them by BS, the size of the pre-/post-test in each case is specified; as can be seen in the figure, the number of students of the Statistics Grade is lower than that in the Computer Science one.

The numbers that appear in the yellow square of Figure 4 correspond to students that never played, that is, that did not participate at all in the activity. With regard to the students that participated, a detailed analysis of participation per *minigame* was performed.

Figure 5 shows the participation in each *minigame*, that is, students from the experimental set that played or not each one. It is interesting to note that the students who played decreased as the academic year advanced. This is probably due to the fact that the work load of the students increases as the course advances.

Considering all the *minigames* and all of the students, the average participation was 66% the first year and 55% the second, with respect to the size of the corresponding experimental sets. 83% of students played at least two games, and 62% three. On average, the students played 3.3 *minigames* each.

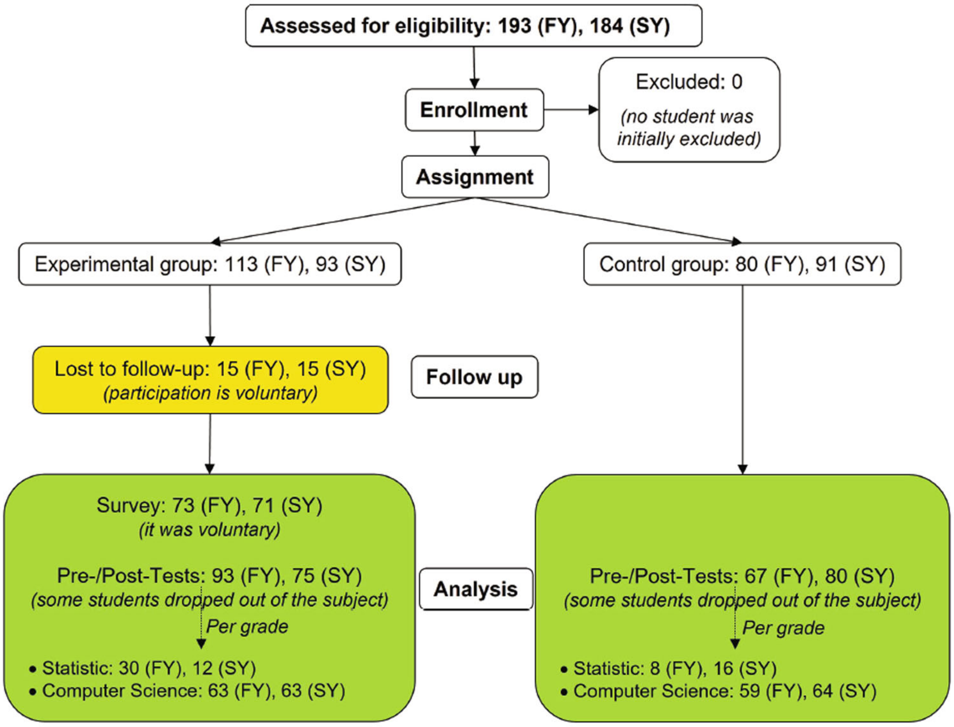
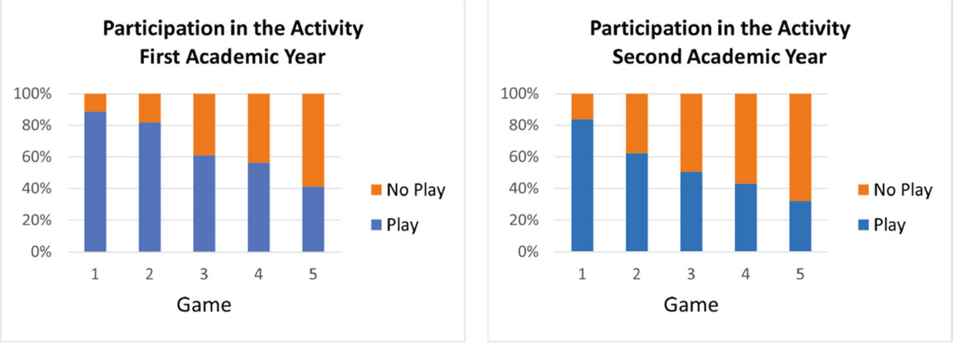


FIGURE 4 | Flow of participants through each stage of the study. "F.Y." stands for First academic Year and "S.Y." for Second academic Year, that is, it is associated with the number of students in the corresponding study stage for each academic year.



(a) Participation. First academic year

(b) Participation. Second academic year

FIGURE 5 | Participation in each minigame of the activity.

Figure 6 shows graphically the participation comparison by BS. The corresponding figures and the statistical analysis can be seen in Table 1. For this analysis, the samples to be compared are created putting "1" if the player plays the game and "0" otherwise. The statistical significance of the differences are also calculated using the *Mann–Whitney–Wilcoxon test*, with H_0: *Computer Science BS and Statistics BS participation samples have the same distribution.*

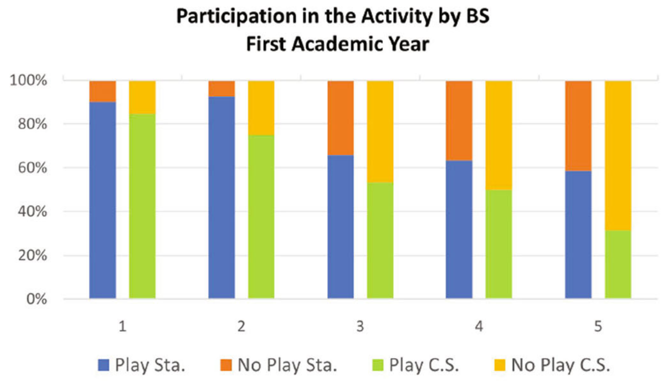
5.1 | Survey Results

Figure 7 shows the survey questions about motivation (question 1), learning (question 2), and user experience or activity opinion (question 3). Here, for a better understanding, the survey questions are shown numbered. In the same way, the answers are also numbered from 1 (most negative opinion) to 5 (most positive), following a Likert-type scale.

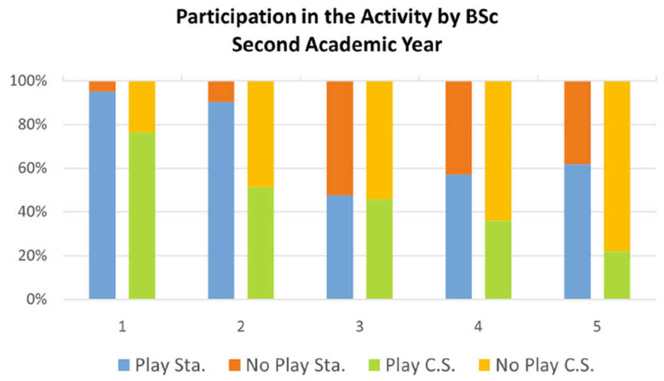
Figure 8 shows the distribution of the answers. The results of the first year (figures (a), (b), and (c)) are in the upper row and in the lower, those of the second year (figures (d), (e), and (f)). Each column shows the results for the questions 1, 2, and 3, respectively.

Table 2 shows the statistical analysis results, following the hypothesis validation proposed in Section 4.2.2. It sets out the probability estimation of improvement (probability of answers 4 or 5) in motivation (**Q1**, related to **PH2**) and learning (**Q2**, related to **PH1**). The probability estimation of satisfaction with the activity (responses 4 or 5 in **Q3**, related with **PH3**) is also shown. This table also shows the statistical significance of these improvements and "likes."

Focusing on the differences between BS, Figure 9 shows the distribution of each answer in each construct. The data



(a) Participation per minigame. First academic year



(b) Participation per minigame. Second academic year

FIGURE 6 | Participation in each minigame of the activity by BS.

TABLE 1 | Participation (%) in each game and the average (*mean* row) by Computer Science students (*C.S.* columns) and Statistics students (*Sta.* columns).

	Academic Year 1			Academic Year 2		
Game	**C.S.**	**Sta.**	***p*-value**	**C.S.**	**Sta.**	***p*-value**
1	85	90	0.39	76	95	**0.001**
2	75	93	**0.02**	52	91	$\mathbf{10^{-5}}$
3	53	66	0.17	46	48	0.25
4	50	63	0.15	36	57	**0.008**
5	32	59	**0.003**	23	62	$\mathbf{10^{-5}}$
Mean	59	75	$\mathbf{10^{-4}}$	47	71	$\mathbf{10^{-11}}$

Note: Those cases where the differences are statistically significant at a 95% confidence level (*p*-value (*U*-test) ≤ 0.05) are boldface emphasized.

1. *The activity fosters the motivation regarding the subject*	2. *I believe that the activity has helped me to understand some concepts about programming in Java*	3. *I liked the activity…*
1. I totally disagree 2. I do not agree 3. I neither agree nor disagree 4. I agree 5. I totally agree	1. Not at all 2. A little 3. Somewhat 4. Quite a lot 5. Very much	1. Not at all 2. A little 3. So so 4. Quite a lot 5. Very much

FIGURE 7 | Questions of the survey.

visualization follows the same lay out as the previous one. The statistical significance analysis of the differences is shown in Table 3, where the mean value of the answers to each question is also shown.

5.2 | Objective Assessment

The hypothesis of the study **PH1**: "The activity boosts students' learning" assessment is again addressed but with objective data. The hypothesis was evaluated by measuring the game's effect on their learning by means of pre- and post-tests and through the final results of the course (Section 4.1).

Table 4 shows **the results of the pre-/post-test** study posed in Section 4.2.1. The results are shown for all the students (*All* rows) and comparing each BS (*Statistics* and *Computer Science* rows) (Figure 4). The mean values of *Dif* and *RelDif* (in % for a better understanding, that is, *RelDif**100) for the experimental and control sets are shown in the corresponding columns; if the value of the column is negative, it means that, on average, the students' grades, either in absolute or relative values are worse in the post-test than in the pre-test, the opposite otherwise.

The difference between these mean values in the experimental and control sets is shown in the column *ExpSet - CtrlSet*; if the value of the column is positive, this means that the performance of the experimental set has been better than that of the control set, the opposite otherwise.

As in [56], the study was extended, eliminating the best and the worst differences that could condition the results. So, the highest and lowest 5% were eliminated, since this was

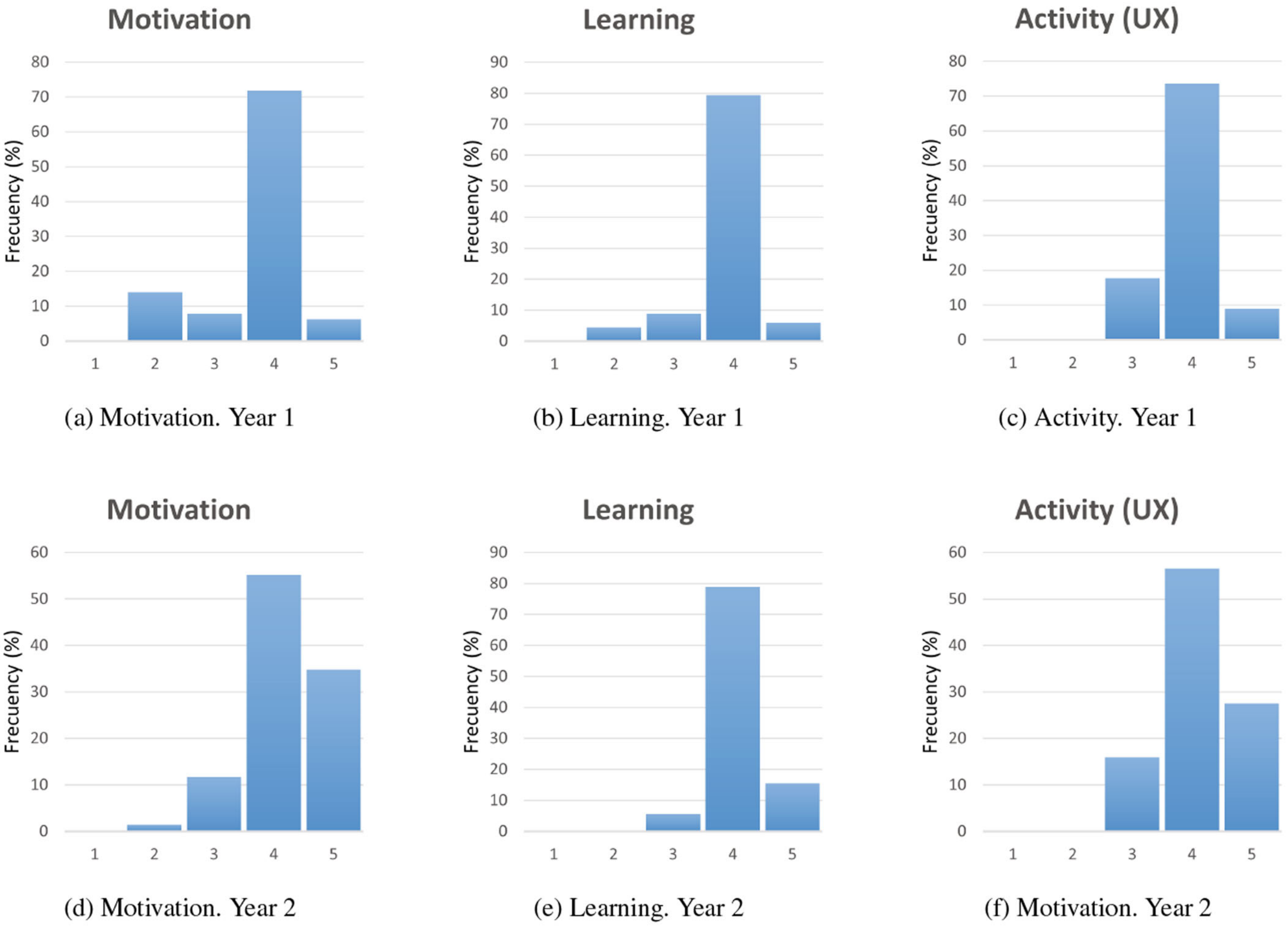


FIGURE 8 | Answer frequency bar plot of questions 1 ((a) and (d)), 2 ((b) and (e)) and 3 ((c) and (f)). In the upper file, the answers of the first year, and in the lower one, those of the second year.

TABLE 2 | Estimation of the Probability of improvement in motivation (Q1) and learning (Q2) or that "the students like the activity" (Q3) (column *P(imp/like)*).

	Academic Year 1				**Academic Year 2**			
	N	**P(imp/like)**	**C.I.**	**Significant?**	***N***	**P(imp/like)**	**C.I.**	**Significant?**
Q1	63	0.94	(0.86,1.00)	Yes	71	0.87	(0.79,1.00)	Yes
Q2	67	0.87	(0.78,1.00)	Yes	71	0.94	(0.88,1.00)	Yes
Q3	79	0.85	(0.77,1.00)	Yes	69	0.84	(0.75,1.00)	Yes

Note: Column *N* shows the number of answers different from the neutral value 3. The statistical significance (Yes/No) of the estimation is shown in the column Significant?, deduced from the confidence interval at 95% (column *C.I.*).

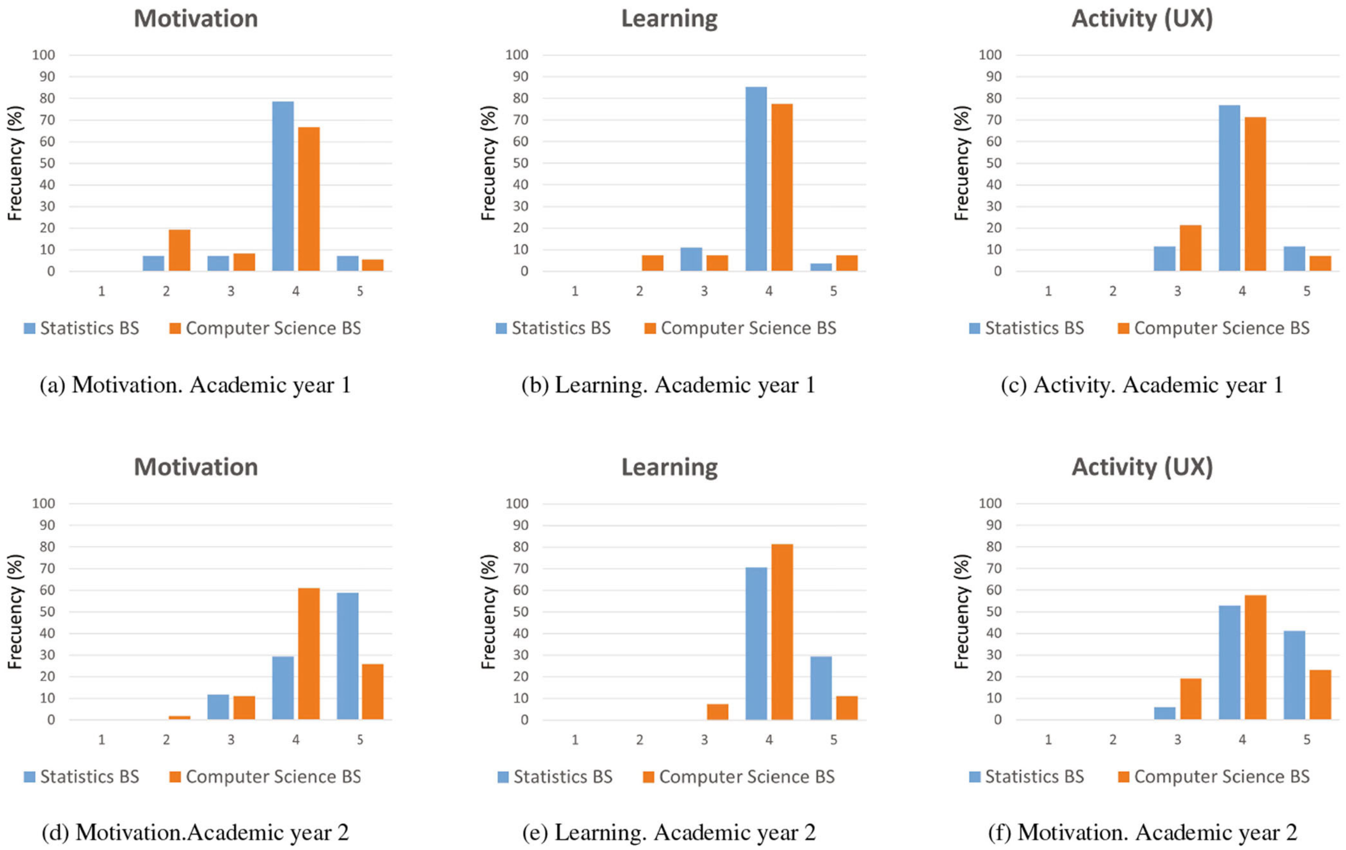
considered a good compromise between removing the extremes, but maintaining an adequate experimental population. Also, for this reason, this study was performed only for the case "All" students. Table 5 shows the results. Analysis of the results based on students' grades or competence can also be seen in other works, for example, refs. [23, 44].

Table 6 shows **the results of the analysis on the final results of the course** with the same segmentation as in the previous section. In the same way, Table 7 shows the analysis when the highest and lowest 5% of the grades had been eliminated, that is, removing the extremes.

6 | Discussion

In this section, the hypotheses and research questions posed are examined in the light of the results shown in the previous section.

Before that, however, it is interesting to provide a brief description of participation in the activity. Keeping in mind that it is voluntary, extracurricular, and that the participants are first year and semester students, a mean participation of over 60% and a mean of 3.3 (out of 5) minigames played can be considered good. However, the lack of research into the use of games as an extracurricular activity does not allow us to



(a) Motivation. Academic year 1 (b) Learning. Academic year 1 (c) Activity. Academic year 1

(d) Motivation.Academic year 2 (e) Learning. Academic year 2 (f) Motivation. Academic year 2

FIGURE 9 | Answer frequency bar plot of questions 1 ((a) and (d)), 2 ((b) and (e)), and 3 ((c) and (f)) for each BS: Statistics (blue bars) and Computer Science (orange bars). In the upper file, the answers of the first year, and in the lower one, those of the second year.

TABLE 3 | Survey answers comparison between Computer Science BS (*Mean C.S.* columns) and Statistics BS (*Mean Sta.* columns) students.

	Academic Year 1			Academic Year 2		
Question	**Mean C.S.**	**Mean Sta.**	***p*-value**	**Mean C.S.**	**Mean Sta.**	***p*-value**
Q1	3.9	4.0	0.26	4.0	4.4	**0.04**
Q2	3.6	3.9	0.20	4.1	4.5	**0.04**
Q3	3.9	4.0	0.92	4.0	4.3	**0.04**

Note: Those cases where the differences are statistically significant at a 95% confidence level (p-value (U-test) ≤ 0.05) are boldface emphasized.

TABLE 4 | Results of the pre-/post-test analysis (Section 4.2.1) per academic year and joining both (*Year* column).

		ExpSet		CtrlSet		ExpSet - CtrlSet			
Year	**Students**	***Dif***	***RelDif* (%)**	***Dif***	***RelDif* (%)**	***Dif***	***p*-value**	***RelDif* (%)**	***p*-value**
First	All	−0.7	−15.0	−1.0	−25.0	**0.3**	0.27	**10**	0.21
	Statistic	−0.2	−1.0	−0.5	−11.4	**0.3**	0.70	**10.4**	0.72
	Computer Science	−0.9	−21.7	−1.0	−26.9	**0.1**	0.65	**5.2**	0.58
Second	All	−0.4	−6.3	−1.0	−16.7	**0.6**	0.06	**10.4**	0.09
	Statistic	0.2	13.0	−0.3	2.4	**0.5**	0.45	**10.6**	0.32
	Computer Science	−0.5	−9.9	−1.2	−21.4	**0.7**	0.07	**11.5**	0.09
Both	All	−0.5	−11.1	−1.0	−20.4	**0.5**	**0.03**	**9.3**	**0.04**
	Statistic	−0.1	3.0	−0.4	−2.2	**0.3**	0.56	**5.2**	0.68
	Computer Science	−0.7	−15.8	−1.1	−24.0	**0.4**	0.08	**8.2**	0.1

Note: Those cases where the experimental set achieves a better performance than the control one are boldface emphasized. To give a wider view of the statistical significance analysis, the following notation has been used: values of p-value > 0.1 (clearly not significant) are in gray, when the p-value ≤ 0.05 (significant difference) is boldface emphasized, and intermediate values $0.1 \geq p$-value > 0.05 (no significant differences at 95%, but low p-value) are in normal text.

TABLE 5 | Results of the pre-/post-test analysis (Section 4.2.1) per academic year and joining both (*Year* column), removing 5% best and worst students. The same notation as in Table 4 is used.

		ExpSet		CtrlSet		ExpSet - CtrlSet			
Year	**Students**	***Dif***	***RelDif* (%)**	***Dif***	***RelDif* (%)**	***Dif***	***p*-value**	***RelDif* (%)**	***p*-value**
First	All	−0.8	−16.3	−1.2	−28.8	**0.4**	0.12	**12.2**	0.07
Second	All	−0.4	−6.1	−1.1	−17.8	**0.7**	**0.02**	**11.7**	**0.04**
Both	All	−0.6	−11.6	−1.2	−22.6	**0.6**	**0.01**	**11**	**0.02**

TABLE 6 | Results of the course final results analysis (Section 4.2.1) per academic year and joining both (*Year* column).

		Grades (Mean)				Approved (%)			
Year	**Students**	**ExpSet**	**CtrlSet**	**Dif.**	***p*-value**	**ExpSet**	**CtrlSet**	**Dif.**	***p*-value**
First	All	4.5	4.0	**0.5**	0.22	47.3	32.8	**14.5**	0.08
	Statistic	5.2	5.4	0.2	0.59	50	50	0.0	1.0
	Computer Science	4.2	3.8	**0.4**	0.41	46.0	30.9	**15.1**	0.09
Second	All	5.1	4.7	**0.4**	0.26	57.3	50	**7.3**	0.22
	Statistic	6.1	5.9	**0.2**	0.66	75.0	68.8	**6.2**	0.74
	Computer Science	4.9	4.4	**0.5**	0.20	54	45.3	**8.7**	0.18
Both	All	4.8	4.4	**0.4**	0.15	51.8	42.7	**9.1**	**0.05**
	Statistic	5.7	5.6	**0.1**	0.54	62.5	60.0	**2.5**	0.67
	Computer Science	4.5	4.0	**0.5**	0.14	50.0	38.7	**11.3**	**0.04**

Note: Those cases where the experimental set achieves a better performance than the control one (*Dif.* column, where *Dif* = Mean (*ExpSet*) − Mean (*CtrlSet*)) are bold face emphasized. To give a wider view of the statistical significance analysis, the following notation has been used: values of p-value > 0.1 (clearly *not significant*) are in gray, when the p-value ≤ 0.05 (significant difference) is bold face emphasized, and intermediate values $0.1 \geq p$-value ≥ 0.05 (no significant differences at 95%, but low p-value) are in normal text.

TABLE 7 | Results of the course final results analysis (Section 4.2.1) per academic year and joining both (*Year* column), removing 5% best and worst grades. The same notation as in Table 6 is used.

		Grades (Mean)			
Year	**Students**	**ExpSet**	**CtrlSet**	**Dif.**	***p*-value**
First	All	4.4	3.7	**0.7**	0.07
Second	All	5.2	4.5	**0.7**	0.1
Both	All	4.8	4.3	**0.5**	0.06

compare these figures with others from the literature. This supports the need for more studies in this context to understand the factors affecting this aspect.

Hypotheses PH2 and PH3, associated to RQ3 and RQ4, respectively, were assessed by means of the survey. The results (Figure 8 and Table 2) show that the students' opinion about the activity is very positive and, futhermore, with a favorable effect on their motivation with respect to the subject.

With regard to **motivation**, in both the first and second academic years, the great majority of the answers have a value of 4, mainly, or 5 (78% in the first year and 88% in the second). In addition, this improvement in the motivation is statistically significant, so the results of the survey confirm PH2, answering RQ3 (Section 1).

Similar results were achieved in the **activity** evaluation. The great majority of students chose mainly answer 4 or 5, consistently in the 2 years (83% in the first year and 85% in the second). As with the results in motivation, this positive opinion about the activity is statistically significant, so the results of the survey also confirm PH3, and answer RQ4.

Comparing the results in both constructs by BS, it is interesting to note that, although the results are good in both, those of the Statistical one are higher in both motivation and activity valuation. These higher results are consistent in both academic years and are even statistically significant in the second.

These findings reinforce the idea that games are entertaining [13], and this turns them into a powerful tool to increase student motivation when used in an educational environment [7, 13, 56, 60–64], even if they are proposed outside the formal curriculum.

Focusing on **learning** (PH1), an essential outcome of the activity, the subjective opinion of the students, collected by means of question 2 of the survey, is very positive, once more choosing mainly answer 4 or 5: 85% in the first year and 94% in the second; these figures being the highest of the three questions of the survey. Therefore, the students think that they have learned through the activity. This subjective opinion is again higher in the students of the Statistic BS in both academic years, being statistically significant in the second.

The subjective feeling shown is important since it reinforces their self-confidence and motivation, which is essential in order to engage the students with the subject and thus improve their learning process [65]; the study was approached objectively by means of the pre-/post-test and the course final results analysis.

The results in the pre-/post-test (Tables 4 and 5) consistently show a better performance of the experimental group, that is, of the students that followed the activity, independently of the students (all the students or only those in the Statistic or Computer Science degrees) and the academic year: the values in the columns *"ExpSet - CtrlSet"* are always positive. In the relative difference, values up to 10% have been achieved in some cases.

Focusing on the statistical analysis, statistically significant differences are achieved when the study extends to the two academic years, so as to have a broader population in the analysis, and when the best and worst cases are removed to eliminate extremes. In the remaining cases, although the differences are not significant, very low p values are achieved in several cases, mostly in the second academic year.

The analysis of the final results of the course shows a similar tendency to that of the pre-/post-test one. Except for the Statistics BS and the first year, the results of the experimental set are consistently better than those of the control. In addition, statistically significant differences were achieved in the analysis of the students who passed. In the same way, when the extreme grades were removed, very low p values were achieved in the differences.

Focusing on the confidence analysis of the differences in the objective study, it should be noted that when this is extended to both courses, thus achieving conditions with large sizes [51] (100 students or bigger), the differences are statistically significant or close to being so (with very low p values). This can be seen in the rows *Both-All* (sizes of the experimental and control sets of 168 and 147, respectively) and *Both-Computer Science* (sizes of the experimental and control sets of 126 and 123, respectively) of Tables 4 and 6.

So, due to all the above, the activity seems to have a positive influence over learning, confirming PH1, and answering the RQ1. This positive influence over learning is in accordance with that shown in previous works [6, 7, 17, 23, 66], but here, it has been extended to extracurricular game-based activities. Another important difference and finding is that this extracurricular approach, allows the use of games to be extended, and thus their positive effect on learning, to all the concepts of an entire course.

Addressing RQ2, the results in Table 4 are positive for students of both BS, without a clear trend for one or the other; although, the students of the BS in Statistics in the experimental set of the second year are the only ones that have a positive value in the "*Dif*" column; which means that, on average, the grades in the post-test (final exam of the subject) are better than in the pre-test. With regard to the final course results (Table 6), except for Statistical BS and the first year, the results are also positive for students of both BS. Since Computer Programming is a "transversal" subject, i.e., it is studied not only in the BS in Computer Science but also in several Engineering or Science degrees, it is interesting to note that the activity seems to foster Computer Programming for both students of Computer Science and external to this discipline, such as the BS in Statistics.

To finish, it is interesting to point out that the results support those previously shown in the literature related to the difficulty of the students when approaching Programming concepts for the first time [12–14]. The results in Table 6 show that, in general, the average grade in the final exam is less than 5.0 (the minimum to pass the exam) and the passing percentage is around 50%, being lower in several cases.

7 | Threats to Validity

Possible threats to validity affecting the results of the study consist of the following: construct, conclusion, and both internal and external threats [67]. These problems are discussed below, looking at how they can be approached so as to be avoided here.

7.1 | Construct Validity

The principal problem with ensuring that what is actually measured is what we wish to measure lies within how the instruments are designed. To overcome the said problem, our survey was both designed and validated by experts in the matter and followed well-defined models [54, 57, 58]. In fact, all the results underwent evaluation. With respect to the objective measurement of the learning, neither the pre-tests nor the post-tests were made in an impromptu manner, as this might have introduced a certain amount of subjectivity. Instead, we used the standard exams and evaluations, which were not connected in any way to the activity, to assess the students' competence in the subject.

7.2 | Internal Validity

Here, the main problems come from the conditions of the data collection and the limitations that arise from the cross-sectional design of the study [68].

The data must be collected as objectively as possible and be as representative as possible in so far as the population under study is concerned so as to avoid bias. An experimental protocol using random sampling was followed to this end for the experimental and control sets, both of which have identical characteristics (same learning activities, assessment, and study materials). The two lecturers of the subject were also similar in experience and knowledge, while the instruments used for testing the activity (survey, pre-tests, and post-tests) were the same for all the students who were, in addition, representative of the study population as a whole.

Concerning implementation, for the sake of objectivity, we concentrated on ensuring the validity of the acquisition measures. The survey was anonymous and the question scale was masked for the students, who were not trained in the use of either the survey or the tests, so the answers were not conditioned by the researchers (*Rosenthal effect*). What is more, the researchers were not generally present during the activities, thus avoiding bias in that sense.

As for the inherent limitations of the design, one of the intentions of the random selection of the sets was to counteract any possible influence of external variables that could be different from the study's independent variables. Even so, in the same way, in order to avoid possible biases in the random selection, the study was extended to two different academic years and BS, thus achieving large experimental and control sets, which minimizes the bias risk [51].

Finally, it should be noted that the researcher who carried out the activity and collected the data was not the same as those who analyzed them, once more in an attempt to avoid any bias. Furthermore, the researchers who analyzed the data were not the lecturers in the subject matter and they did not know the students personally.

7.3 | External Validity

It is not possible at this time to generalize the results because the evaluation was performed in only one university. Nevertheless, due to the nature of the proposal and the participants' profiles, they can be extrapolated for conditions that are similar in so far as education and the socio-cultural sphere are concerned. It may also be possible to extrapolate the results to other socio-cultural and educational conditions as well, since the literature on the topic has shown that the use of serious games can be successful in varying environments, both educationally and culturally (see, for instance refs. [65, 69–71]).

The participants were from two different degree studies, and both had good results. This would seem to show that it is not only an interesting and useful activity for computer science studies.

7.4 | Conclusion Validity

This concerns the aspects that may affect the ability to come to a correct conclusion via the statistical analysis of the data. The main problems in our case were in data collection (sampling, size, representativeness, etc.), being able to use adequate statistical tests and the reliability of the measurements.

Most of these concerns have already been mentioned above, so here, we shall limit our comments to the size of the control and experimental sets. The set that includes all the students and the one with only computer science students both had the minimum number of participants per condition that is considered in ref. [51], but this criterion was not fulfilled for the sets with only students studying the BS in statistics so their results are thus less meaningful.

This study is based on hypotheses and the evaluation was done using survey measurements that have a proven efficacy for measuring the reactions of students. They were also designed with well-defined, validated models using reliable instruments for measuring student competence, including the pre-tests and post-tests.

8 | Limitations

There were two lecturers on the course, something which was unavoidable due to the organization of the teaching responsibilities. This is not ideal for achieving uniformity in the study, but the effect was minimized as both lecturers had similar experience and preparation in the matter. In addition, the materials and tests used were identical for all the students and were created by both lecturers working together.

A second limitation is the nationality of the students, who were all from the same country, with a similar socioeconomic and educational level. Nevertheless, these students can be considered representative of the student population and also similar culturally and educationally to students from other universities.

Another limitation that was beyond the control of the researchers was the time factor. The activity and its assessment were adapted so that the normal course of events was not affected at all, thus avoiding the question of student overload.

9 | Conclusions

In this work, a broad study (315 students, two academic years and two BS) concerning the use of a game-based extracurricular activity is shown. The activity was proposed as support to the Programming Fundamentals learning in higher education courses, encompassing the entire course. The study was based on a Cross-Sectional design, with a between subjects approach, collecting both qualitative and quantitative data. Following this design, the achievement of the activity goals has been evaluated through hypotheses assessment.

An important goal or outcome was to foster student learning. Both qualitative and quantitative data allow us to conclude that the results seem to support the achievement of the said goal. The results of the qualitative assessment were statistically significant. In the quantitative one, the pre-/post-tests and final course results analysis show that the experimental set (students that followed the activity) mainly performed better, independently of the academic year and the BS, with statistically significant differences, or very low p values, in many cases.

Another outcome assessed was motivation and activity experience. The opinion of the students, achieved by means of a survey, has been very positive and statistically significant in both aspects, so it can be concluded that this outcome has also been achieved.

A relevant aspect of the study was to assess the differences in the outcomes between students pursuing a Bachelor in Computer Science, where the programming is a core subject, and those pursuing a Bachelor of Science in Statistics, where programming is an external subject.

From the results, it can be concluded that no differences between Bachelors have been found in the objective learning outcome assessment, as has been pointed out. However, in both the participation in the activity and the survey answers, although the results in both Bachelors have been very positive, those of Statistics have been much higher than those of Computer Science, with statistically significant differences. This poses an interesting future work: extending the activity to other BS.

From all of the above, it can be concluded that the research questions initially posed in the study have been positively answered. The extracurricular game-based activity proposed seems to have a positive influence on learning and is very positively valued by the students. This encourages us to continue in this line of work, extending the activity to other BS, as pointed out, and adding new *minigames* covering new programming concepts.

Acknowledgments

This work was carried out in the Project PID2021-126315OB-I00, which was supported by MCIN/AEI/10.13039/501100011033/FEDER, EU.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. L. Nagowah and D. Cuniah, "Code-Venture: A Mobile Serious Game for Introductory Programming," in Software Engineering for Games in Serious Contexts: Theories, Methods, Tools, and Experiences (Springer Nature Switzerland, 2023), 155–187, https://doi.org/10.1007/978-3-031-33338-5_8.

2. Z. J. Fitz-Walter, "Achievement Unlocked: Investigating the Design of Effective Gamification Experiences for Mobile Applications and Devices" (PhD thesis, Queensland University of Technology, 2015).

3. Z. Sabri, A. Remaida, A. Benyoussef, et al., " Engineering Education Using Serious Games: The Current State, Trends and Future Using a Bibliometric Study," in Automatic Control and Emerging Technologies, eds. H. El Fadil, W. Zhang (Springer Nature Singapore, 2024), 113–127.

4. D. R. Michael and S. L. Chen, *Serious Games: Games that Educate, Train, and Inform* (Muska & Lipman/Premier-Trade, 2005).

5. E. Klopfer, *Augmented Learning: Research and Design of Mobile Educational Games* (MIT Press, 2008).

6. L.-H. Wang, B. Chen, G.-J. Hwang, J.-Q. Guan, and Y.-Q. Wang, "Effects of Digital Game-Based Stem Education on Students' Learning Achievement: A Meta-Analysis," *International Journal of STEM Education* 9, no. 1 (2022): 1–13.

7. J. Zeng, S. Parks, and J. Shang, "To Learn Scientifically, Effectively, and Enjoyably: A Review of Educational Games," *Human Behavior and Emerging Technologies* 2, no. 2 (2020): 186–195.

8. M. Csikszentmihalyi, *Flow: The Psychology of Optimal Experience* (Harper & Row, 1990).

9. M. Csikszentmihalyi, *Beyond Boredom and Anxiety* (Jossey-Bass, 2000).

10. E. L. Deci and R. M. Ryan, "Self-Determination Theory: When Mind Mediates Behavior," *Journal of Mind and Behavior* 1 (1980): 33–43, https://www.jstor.org/stable/43852807.

11. R. M. Ryan and E. L. Deci, "Self-Determination Theory and the Facilitation of Intrinsic Motivation, Social Development, and Well-Being," *American Psychologist* 55, no. 1 (2000): 68.

12. N. Humble, P. Mozelius, and L. Sallvin, " Need-Based Game Design: Developing a CT Game Based on Educational Needs," in Proceedings of the 17th European Conference on Game-Based Learning: ECGBL 2023 (2023), 269–277, https://www.scopus.com/inward/record.uri?eid=2-s2.0-85179007194&partnerID=40&md5=f564122186b99c034608fa612f467d7c.

13. D. Zhao, C. H. Muntean, A. E. Chis, G. Rozinaj, and G.-M. Muntean, "Game-Based Learning: Enhancing Student Experience, Knowledge Gain, and Usability in Higher Education Programming Courses," *IEEE Transactions on Education* 65, no. 4 (2022): 502–513.

14. K. A.-M. Sarpong, J. K. Arthur, and P. Y. O. Amoako, "Causes of Failure of Students in Computer Programming Courses: The Teacher-Learner Perspective," *International Journal of Computer Applications* 77, no. 12 (2013): 27–32.

15. C. Montuori, F. Gambarota, G. Altoe, and B. Arfe, "The Cognitive Effects of Computational Thinking: A Systematic Review and Meta-Analytic Study," *Computers & Education* 210 (2024): 104961, https://www.sciencedirect.com/science/article/pii/S0360131523002385.

16. M. Videnovik, T. Vold, L. Kiønig, A. M. Bogdanova, and V. Trajkovik, "Game-Based Learning in Computer Science Education: A Scoping Literature Review," *International Journal of STEM Education* 10 (2023): 54.

17. A. Theodoropoulos and G. Lepouras, "Digital Game-Based Learning and Computational Thinking in P-12 Education: A Systematic Literature Review on Playing Games for Learning Programming," in *Handbook of Research on Tools for Teaching Computational Thinking in p-12 Education* (IGI Global, 2020), 159–183.

18. I. Garcia, C. Pacheco, F. Méndez, and J. A. Calvo-Manzano, "The Effects of Game-Based Learning in the Acquisition of "Soft Skills" on Undergraduate Software Engineering Courses: A Systematic Literature Review," *Computer Applications in Engineering Education* 28, no. 5 (2020): 1327–1354.

19. F. Zhang and D. Kaufman, "Using Video Games in Computer Science Education," *European Scientific Journal, ESJ* 10 (2014): 37–52, https://api.semanticscholar.org/CorpusID:116604312.

20. G. Charness, U. Gneezy, and M. A. Kuhn, "Experimental Methods: Between-Subject and Within-Subject Design," *Journal of Economic Behavior & Organization* 81, no. 1 (2012): 1–8.

21. M. Appelbaum, H. Cooper, R. B. Kline, E. Mayo, A. M. Nezu, and S. M. Rao, "Journal Article Reporting Standards for Quantitative Research in Psychology: The APA Publications and Communications Board Task Force Report," *American Psychologist* 73, no. 1 (2018): 3–25.

22. C. Vicuña, What are Europe's most needed software roles and skills? Trends to look out for (2022), accessed November 16, 2023, https://digital-skills-jobs.europa.eu/en/latest/news/what-are-europes-most-needed-software-roles-and-skills-trends-look-out.

23. D. Zhao, C. H. Muntean, A. E. Chis, and G.-M. Muntean, "Learner Attitude, Educational Background, and Gender Influence on Knowledge Gain in a Serious Games-Enhanced Programming Course," *IEEE Transactions on Education* 64, no. 3 (2021): 308–316.

24. M. W. Ohland, S. D. Sheppard, G. Lichtenstein, O. Eris, D. Chachra, and R. A. Layton, "Persistence, Engagement, and Migration in Engineering Programs," *Journal of Engineering Education* 97, no. 3 (2008): 259–278.

25. S. AlJanah, P. S. Teh, J. Y. Tay, O. Aiyenitaju, and R. Nawaz, " Minecraft as a Tool to Enhance Engagement in Higher Education," in Research and Innovation Forum 2022, eds. A. Visvizi, O. Troisi and M. Grimaldi (Springer International Publishing, 2023), 465–476.

26. A. Giannakoulas and S. Xinogalos, "Studying the Effects of Educational Games on Cultivating Computational Thinking Skills to Primary

School Students: A Systematic Literature Review," *Journal of Computers in Education* 11 (2023): 95–108.

27. M. A. Miljanovic and J. S. Bradbury, "GidgetML: An Adaptive Serious Game for Enhancing First Year Programming Labs," in Proceedings of the ACM/IEEE 42nd International Conference on Software Engineering: Software Engineering Education and Training (2020), 184–192.

28. A. Areizaga Blanco and H. Engstrom, "Patterns in Mainstream Programming Games," *International Journal of Serious Games* 7, no. 1 (2020): 97–126, https://journal.seriousgamessociety.org/index.php/IJSG/article/view/335.

29. J. P. d. Silva and I. F. Silveira, "A Systematic Review on Open Educational Games for Programming Learning and Teaching," *International Journal of Emerging Technologies in Learning (IJET)* 15, no. 9 (2020): 156–172, https://online-journals.org/index.php/i-jet/article/view/12437.

30. J. Ekman, Comparing Engagement of VR and non-VR Games for Programming Learning Purposes (Master's thesis, Stockholm University, Department of Computer and Systems Sciences, 2023).

31. R. Sims, N. Rutherford, P. Sukumaran, N. Yotov, T. Smith, and A. Karnik, "Logibot: Investigating engagement and development of computational thinking through virtual reality," in 2021 7th International Conference of the Immersive Learning Research Network (iLRN) (2021), 1–5.

32. R. J. Segura, F. J. del Pino, C. J. Ogayar, and A. J. Rueda, "VR-OCKS: A Virtual Reality Game for Learning the Basic Concepts of Programming," *Computer Applications in Engineering Education* 28, no. 1 (2020): 31–41, https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.22172.

33. P. Toukiloglou and S. Xinogalos, "Adaptive Support with Working Examples in Serious Games about Programming," *Journal of Educational Computing Research* 61, no. 4 (2023): 766–789, https://doi.org/10.1177/07356331231151393.

34. D. Dicheva and A. Hodge, " Active learning through game play in a data structures course," in Proceedings of the 49th ACM Technical Symposium on Computer Science Education (2018), 834–839.

35. B. Law, " Introducing Novice Programmers to Functions and Recursion Using Computer Games," in ECGBL 2018 12th European Conference on Game-Based Learning, Proceedings of the European Conference on Games-based Learning, ed. M. Ciussi (Academic Conferences and Publishing International Limited, 2018), 325–334.

36. T. Hainey, G. Baxter, J. Black, et al., "Serious Games as Innovative Formative Assessment Tools for Programming in Higher Education," in Proceedings of the 16th European Conference on Games Based Learning, Vol. 16, 2022, 253–261.

37. T. Petrusic and D. Novak, "A 16-week School-Based Intervention Improves Physical Fitness in Slovenian Children: A Randomized Controlled Trial," *Frontiers in Physiology* 15 (2024): 1–10, https://www.frontiersin.org/journals/physiology/articles/10.3389/fphys.2024.1311046.

38. S. Lopez-Serrano, A. Ruiz-Ariza, M. D. L. Torre-Cruz, and E. J. Martinez-Lopez, "Improving Cognition in School Children and Adolescents through Exergames. A Systematic Review and Practical Guidel," *South African Journal of Education* 14, no. 1 (2023): 1–19, https://www.ajol.info/index.php/saje/article/view/205732.

39. E. M. Melvin, B. Bowles, and A. Steele, " Using Escape Rooms to Apply Team Building and Leadership Skills in an Engineering Leadership Development Program: A Work in Progress," in 2023 ASEE Annual Conference & Exposition (ASEE Conferences, 2023), https://doi.org/10.18260/1-2-44575.; https://peer.asee.org/44575.

40. D. Ditter, I. Rubner, S. Lukas, and D. Weiser, " Science4exit Project - Experimental Escape Games With Digital Enrichment in an Extracurricular Learning Venue," in Proceedings of the 17th European Conference on Games Based Learning, Vol. 17 (2023), 154–261.

41. L. Bosova, *Program, Learn and Play: A Course of Extracurricular Activities in Scratch Programming for Students in Grades 3–6* (Springer International Publishing, 2023), 113–126, https://doi.org/10.1007/978-3-031-21970-2_8.

42. C. Gresse von Wangenheim, J. C. R. Hauck, F. S. Pacheco, and M. F. Bertonceli Bueno, "Visual Tools for Teaching Machine Learning in k-12: A Ten-Year Systematic Mappingl," *Education and Information Technologies* 26, no. 5 (2021): 5733–5778.

43. J. Perrin, A. Meeus, J. Broseus, et al., "A Serious Game About Hematology for Health Care Workers (Super Hemo): Development and Validation Study," *JMIR Serious Games* 11 (2023): e40350, https://games.jmir.org/2023/1/e40350.

44. R. S. Nagovitsyn, E. A. Berezhnykh, A. E. Popovic, and O. V. Srebrodolsky, "Formation of Legal Competence of Future Bachelors of Psychological and Pedagogical Education," *International Journal of Emerging Technologies in Learning (IJET)* 16, no. 1 (2021): 188–204, https://online-journals.org/index.php/i-jet/article/view/17591.

45. P. M. Johnson, C. Moore, P. Leong, and S. Paek, " Radgrad: Removing the "extra" from Extracurricular to Improve Student Engagement, Diversity, and Retention," in Proceedings of the 51st ACM Technical Symposium on Computer Science Education, SIGCSE '20 (Association for Computing Machinery, 2020), 746–751, https://doi.org/10.1145/3328778.3366793.

46. J. Schell, *The Art of Game Design* (A K Peters/CRC Press, 2019). 3rd ed.

47. I. Anton-Solanas, B. Rodriguez-Roca, F. Urcola-Pardo, et al., "An Evaluation of Undergraduate Student Nurses' Gameful Experience Whilst Playing a Digital Escape Room as Part of a First Year Module: A Cross-Sectional Study," *Nurse Education Today* 118 (2022): 105527, https://www.sciencedirect.com/science/article/pii/S0260691722002635.

48. K. Bovermann and T. J. Bastiaens, "Towards a Motivational Design? Connecting Gamification User Types and Online Learning Activities," *Research and Practice in Technology Enhanced Learning* 15 (2020): 1.

49. D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating Training Programs* (Berrett-Koehler Publishers, 2006). 3rd ed.

50. E. A. Boyle, T. Hainey, T. M. Connolly, et al., "An Update to the Systematic Literature Review of Empirical Evidence of the Impacts and Outcomes of Computer Games and Serious Games," *Computers & Education* 94 (2016): 178–192.

51. A. All, E. P. N. Castellar, and J. V. Looy, "Assessing the Effectiveness of Digital Game-Based Learning: Best Practices," *Computers & Education* 92–93 (2016): 90–103.

52. D. López-Fernández, A. Gordillo, P. P. Alarcón, and E. Tovar, "Comparing Traditional Teaching and Game-Based Learning Using Teacher-Authored Games on Computer Science Education," *IEEE Transactions on Education* 64, no. 4 (2021): 367–373.

53. G. Petri, C. Gresse von Wangenheim, and A. F. Borgatto, " MEEGA +, Systematic Model to Evaluate Educational Games," in Encyclopedia of Computer Graphics and Games (Springer International Publishing, 2017), 1–7, https://doi.org/10.1007/978-3-319-08234-9_214-1.

54. R. Savi, C. G. von Wangenheim, and A. F. Borgatto, " A Model for The Evaluation of Educational Games for Teaching Software Engineering," in 25th Brazilian Symposium on Software Engineering (2011), 194–203.

55. R. F. DeVellis, *Scale Development: Theory and Applications* (Thousand Oaks, California, USA: Sage Publications, 2003). 2nd ed.

56. A. Pisabarro-Marron, C. Vivaracho-Pascual, E. Manso-Martinez, and S. Arias-Herguedas, "A Proposal for an Immersive Scavenger Hunt-Based Serious Game in Higher Education," *IEEE Transactions on Education* 67, no. 1 (2024): 131–142.

57. F.-L. Fu, R.-C. Su, and S.-C. Yu, "Egameflow: A Scale to Measure Learners' Enjoyment of E-Learning Games," *Computers & Education* 52, no. 1 (2009): 101–112.

58. E. D. Mekler, F. Brühlmann, A. N. Tuch, and K. Opwis, "Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance," *Computers in Human Behavior* 71 (2017): 525–534.

59. J. Cohen, *Statistical Power Analysis for the Behavioral Sciences* (New York, USA: LEA Publishers, 1988). 2nd ed.

60. S. Uluskan, "Eduperes and Gameres: A Smart Interactive Education Platform and Digital Gamification for Basic Circuit Analysis," *Computer Applications in Engineering Education* 32, no. 4 (2024): e22754, https://onlinelibrary.wiley.com/doi/abs/10.1002/cae.22754.

61. A. Min, H. Min, and S. Kim, "Effectiveness of Serious Games in Nurse Education: A Systematic Review," *Nurse Education Today* 108 (2022): 105178.

62. S. Subhash and E. A. Cudney, "Gamified Learning in Higher Education: A Systematic Review of the Literature," *Computers in Human Behavior* 87 (2018): 192–206.

63. C. A. Bodnar, D. Anastasio, J. A. Enszer, and D. D. Burkey, "Engineers at Play: Games as Teachingtools for Undergraduate Engineering Students," *Journal of Engineering Education* 105, no. 1 (2016): 147–200.

64. M. Qian and K. R. Clark, "Game-Based Learning and 21st Century Skills: A Review of Recent Research," *Computers in Human Behavior* 63 (2016): 50–58.

65. H. Almas, F. Pinkow, and F. Giæver, "Reimagining How to Understand Learning Game Experiences: A Qualitative and Exploratory Case Study," *International Research Journal of Engineering, IT & Scientific Research* 10 (2023): 14.

66. I. n. Aldalur, "Gamifying Software Engineering Subject to Enhance the Quality of Knowledge," *Software: Practice and Experience* 54, no. 12 (2024): 1–17, https://onlinelibrary.wiley.com/doi/abs/10.1002/spe.3339.

67. C. Wohlin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell, and A. Wesslen, *Experimentation in Software Engineering* (Springer-Verlag, 2012). 1st ed.

68. M. S. Setia, "Methodology Series Module 3: Cross-Sectional Studies," *Indian Journal of Dermatology* 61, no. 3 (2016): 261–264.

69. D. Alajaji and A. A. Alshwiah, "Effect of Combining Gamification and a Scavenger Hunt on Pre-Service Teachers' Perceptions and Achievement," *Journal of Information Technology Education: Research* 20 (2021): 283–308.

70. A. Q. Khan, " Serious Games in Science Education: A Systematic Mapping," in ECGBL 2021 15th European Conference on Game-Based Learning (2021), 412–418, XVII.

71. C. Zaror, R. Mariño, and C. Atala-Acevedo, "Current State of Serious Games in Dentistry: A Scoping Review," *Games for Health Journal: Research, Development, and Clinical Applications* 10, no. 2 (2021): 95–108.

Appendix A

Course Content and Study Schedule

Table A1 shows the content of the course and the study activities schedule.

TABLE A1 | Course, game-based and study activities schedule.

Week/s	Course content	Activity
1	Programming environment	
2–3	Lexical, Variables, and Basic Data Types	
4	Methods	
5	Conditional Control Structures	midterm exam (pre-test)
6	Conditional Control Structures	"Data Drop" *minigame* is open
7	Iteration Control Structures	
8	Iteration Control Structures	"Mars Miners" *minigame* is open
9	Recursion	
10–11	Vectors and Multidimensional Arrays	"Recursive Party" *minigame* is open
12	Heterogeneous Data Structures	"Pointed" *minigame* is open
12	Heterogeneous Data Structures	"Stacked" *minigame* is open
14	Files	
15	Dynamic Data Structures	Survey
School Period End. Examination period		
19		Final exam (post-test)