

# A Proposal for an Immersive Scavenger Hunt-Based Serious Game in Higher Education

Alma Pisabarro-Marron, Carlos Vivaracho-Pascual<sup>1b</sup>, Esperanza Manso-Martinez, and Silvia Arias-Herguedas

**Abstract—Contribution:** A successful activity based on the scavenger hunt (SH) game is presented here. Although “serious game” in education now seems synonymous with videogame, the effectiveness of hands-on traditional games to increase student performance that, besides, they also like is defended and proved. The proposal is not focused, as is usual, on a single aspect of the educational environment, but on integrating behavioral and affective aspects into the learning process.

**Background:** The literature analysis shows the predominance of serious games based on videogames in education, perhaps due to the lack of objective evidence concerning the influence of traditional game alternatives, such as SHs, on students’ attitude/learning. This objective evidence is addressed with a proposal to motivate and integrate the students, making them more participative and thus positively affect their learning.

**Intended Outcomes:** The activity increases motivation (behavioral outcome) and socialization (affective outcome), boosting learning (competence outcome); besides, students like the activity.

**Application Design:** The necessity for a different instructional strategy came from the lack of commitment by the students in the first year of Computer Science Engineering. So a game (they are entertaining and powerful tools to increase motivation) that takes the students outside their normal working environment (classroom and laboratory) was designed. The study follows a cross sectional design with experimental and control sets randomly created, and sizes of 106 and 98 students, respectively.

**Findings:** Highly satisfactory and statistically significant results were achieved: their attitude in class and personal study was more active (motivation), new relationships were created (socialization), they obtained better marks (learning) and enjoyed the activity (user experience), even though it was nondigital.

**Index Terms—**Computer science (CS), educational research, engagement, motivation, serious games.

## I. INTRODUCTION

THE ACTIVITY presented and evaluated here was initially proposed as an alternative to solve a problem. In our case, the lack of commitment on the part of many of the students in Fundamentals of Programming. The students were a heterogeneous group of first year students from three

Manuscript received 26 May 2022; revised 24 April 2023; accepted 23 October 2023. (Corresponding author: Carlos Vivaracho-Pascual.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Ethical Committee of the University of Valladolid under Application No. UVA/10/2023.

The authors are with the Computer Science Department, University of Valladolid, 47011 Valladolid, Spain (e-mail: alma@infor.uva.es; cevp@infor.uva.es; espe@infor.uva.es; silvia@infor.uva.es).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TE.2023.3330764>, provided by the authors.

Digital Object Identifier 10.1109/TE.2023.3330764

different degrees (Computer Engineering, Statistics, and the double degree in Statistics and Computing), with different levels of motivation and commitment.

For the design of the activity, and following the terminology of the Instructional Design Model ADDIE (Analysis → Design → Development → Implementation → Evaluation) [1], [2], after analyzing the teaching context (target audience), it was considered what new instructional strategy could be designed to deal with the problem, as this disparity in levels of participation made it very difficult to advance in the subject. Something to improve student motivation was needed. The answer was to design a game. Games are entertaining, and this turns them into a powerful tool to increase student motivation when used in an educational environment, including university [3], [4], [5], [6], [7]. In addition, serious games are based on educational theories which suggest that learning is more effective when it is active, experimental and based on problems, aligning well with different learning theories, as, for example, *constructivism*, *Activity theory*, *flow theory*, or *Experiential and generative learning* [7].

Although the predominant tendency to use digital games (Section II) fits students in computing like a glove, given the problem as set out above, a completely different approach was decided: a scavenger hunt (SH) type game that would take the students out of their habitual surroundings (classroom and laboratory). The participants had to physically move around the School to progress in the game. This made the game more attractive (as seen in the evaluation results), and allowed them to get to know the School and all its members better, including their future teachers. In order to make the experience more immersive and give it coherence and cohesion, the activity was set within an attractive story well-known to the students; Harry Potter was selected for the first year, using other sagas (e.g., Game of Thrones, the Hunger Games, or the Avengers) in following editions. The inclusion of narrative has been shown to foster immersion in video games-based educational serious games [8], but, to the best of our knowledge, it has not been included in those based on SH.

The first year, the activity proved to be highly successful. A substantial improvement in the students’ participation and motivation was perceived. The subject had two groups, and the activity was initially done for only one, but, at the students’ request, the game had to be repeated for the other. Such a positive response made us continue with the activity in the following years, broadening its scope and delving into its outcomes, since it was seen that not only was it possible to increase the students’ motivation and participation but also to:

- 1) help them in the process of socialization
- 2) positively influence their learning, making the educational process a little more attractive.

Although motivation is a subjective process of each person, it is the motor that allows people to progress satisfactorily in all their activities [9]. In the educational field, motivation is the state of mind that students manifest in their teaching-learning process, that is, the interest they have in creating their own learning by building their own knowledge [10]. So, based on *self-determination theory*, “motivation and learning are linked, with motivation depending on the intrinsic needs of competence, autonomy, and relatedness” [11], which also mean that an affective component, such as socialization, is also important in the learning process of the students. These are the theoretical bases of gamification in this study.

This article describes the activity and shows the systematic and empirical study performed for its evaluation with the goal of achieving objective evidence of the proposed outcomes (goals). To report this quantitative research, this article is structured following the JARS-Quant (Journal Article Reporting Standards for Quantitative Research) template [12].

The main differences of the study presented here with regard to those in the literature about serious games, both in general and based on SH in particular, are (these being the main contributions of this work) as follows.

- 1) It is not a videogame, as is usual (Section II). The utility of serious games based on traditional games as an interesting alternative, which the students also like, is proposed and demonstrated (Section VII).
- 2) The majority of nondigital proposals [6], [13], [14], are based on card or building games in which the player remains in the same place, usually around a table. The students are taken out of their normal working environment.
- 3) One of the main drawbacks regarding the use of SH in education is the lack of objective evidence about its outcomes [15], [16]. Most studies discussed the game/activity evaluation more than its impact on the students [17] (Section II). Here, special attention has been paid to the evaluation of the activity outcomes to obtain empirical evidence of improvements in motivation, socialization and learning, which correspond to the three main outcomes identified when games are integrated into the learning process: a) behavioral; b) affective; and c) cognitive [18]. The game (“user experience”) has also been evaluated, comparing the results with the students’ gaming profile, following the best practices in [19].
- 4) It is not an average serious game or SH that can be found in the literature, which in general focus on a single aspect of the educational environment, as will be seen in Section II. The proposal is different and focuses on a wider vision of student learning that integrates behavioral and affective aspects into the learning process. The aim is to motivate and integrate the students, making them more participative, and thus positively affect their learning, i.e., not only to engage them with the activity (extrinsic motivation) but that the students should remain

engaged with the subject after the activity (intrinsic motivation [20]).

This last characteristic makes the game adaptable to any subject or material, since encouraging students to participate more proactively and making them more curious about the concepts is a common objective in any learning process.

All procedures performed in this study were in accordance with ethical and local legal regulations, following the guidelines included in document Ref. UVA/10/2023, approved by the institutional review board of the University. Before the study, all participants were informed of its purpose and it was guaranteed that their participation or withdrawal would incur neither reward nor punishment.

## II. RELATED WORKS

This section focuses on the state of the art in the use of SHs in education, related with other serious games alternatives, to be compared with the proposal here. For this, a formal search protocol has been followed to find the most recent and relevant related works. The most important scientific databases (IEEE Xplore, WOS (Web Of Science) and Scopus) were used as *information sources*. The *search terms* were *serious AND games AND education AND review* for works concerning serious games in general (this was focused on reviews given the great amount of related works) and (*scavenger OR treasure*) *AND hunt* for works concerning the use of SH. The *Selection criteria* were relevance and publication year: only recent works were considered, in general. Despite the formal search, it is important to note the possibility of the literature not considered due to *publication bias, fugitive literature, or gray literature*, for example.

The goal was to answer the following questions.

- 1) For serious games in education: What types of serious games are proposed? Can examples of SH be found?
- 2) For SH use: What is their application field? What are their goals? How are their outcomes assessed?

Almost 500 works were analyzed in both searches, using the title and abstract to perform the final selection. 59 works concerning reviews of serious games in education and 71 related to the use of SH as serious game were selected for analysis. Only the most relevant references are included here.

The first conclusion from the reviews of serious games in education is the growth in their use as a new teaching/learning methodology to better connect with the students. Many reviews were found, with both a general character [4], [6], [18], [21], [22], [23] and focusing on particular applications or game aspects [3], [24], [25], [26], [27], [28].

Another interesting tendency is the use of digital games, since the great majority of the works and reviews focus on them. Nondigital games were found in [6], [13], and [14], but very few regarding digital ones, and none in the more recent literature. Since today’s students have grown up in the digital era, this tendency is logical. However, the use of nondigital games can be an interesting proposal, precisely because it is different. In addition, if instead of sitting around a table or at a computer, they move around the School/Campus, the game can catch their attention and interest more than a digital one.

Focusing on the use of SH, many works have been found in the scientific literature, but none with the approach and assessment shown here, as seen below.

The areas of application as a serious game are varied. In fact, only 39 of the 71 analyzed concerned education.

Regarding the outcomes, the main goal pursued with the use of a SH is to improve knowledge of the environment. Focusing on education, two main aspects have been addressed: to improve the campus library knowledge (six works, e.g., [15]) and to familiarize students with a new academic environment and its resources (nine works, e.g., [29] and [30]). Papers not concerning the educational environment can also be found, e.g., [31] and [32]. The second most important goal addressed is to learn or promote the learning of a concrete concept or lesson (e.g., [33], [34], [35], and [36]), normally including questions related to these concepts in the SH tasks. The SH is also used as a socialization tool, usually for first year students [29], [30], [37], but also to improve collaborative skills [38], [39]. Motivation, one of the main outcomes here and, in general, of serious games in education [4], [17], has only been found in two works, but solely in [34] is it used in the same way as here, namely, to modify student attitudes in order to foster their learning, but with a different approach; in [40], the motivation is with respect to the activity not learning.

There are two works found in the literature that have similarities with the proposal shown here [34], [41], but with a different approach, where the game is not used to learn a specific concept, but to boost the learning of the subject. Both works have the same educational environment: students of a Degree in Primary Education (preservice teachers in early childhood education) and the subject is related to gamification. Furthermore, in both, one of the goals with the SH is for students to become familiar with a gamification platform, Quizizz and WebQuest, respectively. Therefore, the SH is more a “laboratory activity” (in engineering terminology) to apply or practice the theoretical lesson than a way to motivate the students, i.e., in these works, the motivation can come more from the practical application of the theory seen in the classroom and their use later in their careers than from the game itself, as in the present study. It is impossible to distinguish both sources of motivation in the results shown in the said works.

Another interesting consideration is the type of implementation: virtual versus nonvirtual. Several examples of completely virtual SH can be found in recent years, usually based on virtual reality [42], [43]. However, the most popular type is one where the student must move through the campus/school/city, normally using the GPS of the mobile, to locate the “clues” in the game, e.g., [44] and [45]. Many software applications were also found, both for mobiles (e.g., [44], [46], and [40]) and platforms for general use (e.g., Quizizz [34] and WebQuest [35], [41]). In [33], where the students can play outdoors or at home, the former performed significantly better. This is the option here, that the students must move around during the game. Reference [43] is the only work found with a comparative study between playing a traditional SH versus a virtual one; in all of the areas analyzed,

except one (students’ perception that librarians and staff want to help them), both versions performed the same. Here, a traditional version has been followed, while also showing that the students liked it.

One of the more important aspects in any instructional strategy is its evaluation, i.e., to assess if the goals posed have been achieved or not. This has been analyzed in depth in the literature review. The most important conclusion is that we agree with [15] and [16] about the lack of objective evidence concerning the outcomes in the use of SH in education. From the 71 articles analyzed only 31 include any kind of evaluation, of which only 16 focus on the effect on the students’ attitude/learning. The remaining ones focus on the game (or software application) evaluation. All 16 works that study the impact of the activity on the students have one or more of the following important drawbacks: 1) they do not follow an experimental or empirical protocol, generally because a control set is not used; 2) the results are not statistically analyzed to calculate if they are significant or not; and 3) related to the previous item, a very limited number of students participated in the study, generally less than 50.

### III. DESCRIPTION OF THE ACTIVITY

The activity followed the traditional SH type game format, which entails a sequence of tasks, each one consisting of two parts.

- 1) The first is a moderately complex calculation that to solve it the students have to implement a program. The result is a number of approximately 10 digits or a word. This “key” must be presented at the following collection point in order to receive the envelope with the next task.
- 2) The second is a test of three questions with three possible answers each: A, B or C. The correct sequence of answers indicates where they have to pick up the envelope with the next task. The students have a directory with all the possible combinations of the characters A, B and C which are associated to an office or concrete location within the school building.

The number of tasks was fixed at three and each one was more difficult than the previous. The envelope with the first task was handed to them at the start of the game.

Although the participants could move freely around the school [Fig. 1(a)], their principal area of work was in the School common room [Fig. 1(c)]. Before the students’ arrival, the room was prepared with a table for each group, identified by a card with their name and a ribbon with their color. On each table there was: a copy of the rules, the directory and neckerchiefs with the identifying color for each member of the group [Fig. 1(b)]; all the participants had to wear this in a visible place for the entire duration of the activity.

The name of the activity always followed the chosen narrative thread, together with some informatics terminology. For example, it was called “Progravers” when the storyline was based on the Avengers of Marvel.

The tests were also related to the narrative thread; e.g., in Progravers, they were related to the battle against Thanos: The Chitauri Invasion, The Enigma of Vormir and The Attack



Fig. 1. Pictures of the activity. (a) Participants moving around the school. (b) Team with their identification neckerchief. (c) Principal area of work. (d) Final goal objects. (e) Medals for the participants.

of Thanos. The heading of each task, where the problem to be solved was set out, also followed the main theme. This material is available in: <https://greidi.infor.uva.es/material.php>.

The players' final goal consisted in recuperating a physical object (e.g., the gauntlet with the Infinity Gems in Progravengers) [Fig. 1(d)]. This object was unique and was in a public place in the principal game room. To avoid problems in order to determine who the winner was, before handing over the last key, the team had to collect the object and give it to the lecturer together with the key. If it was incorrect, the object had to be returned to its place. As the object was unique, the "referee" had no doubt about which key to check first. The carrier of the object had priority.

The activity was presented with an explanation that introduced the students to the narrative thread. For the Progravengers it was "*The villain Thanos already possesses most of the infinity gems. The Soul gem is the only one still beyond his reach and is guarded by monks in the Himalayas. Our mission is to stop Thanos from completing the collection with the gauntlet of infinity, which would give him supreme power over all creatures in the universe. The war of infinity is about to begin and only those men and women of the greatest skill and courage have been selected to participate in the fight. Congratulations soldier! To be able to join the ranks you must add your name to the list below:*"

The students could choose the avatar they wanted to use. In Progravengers, some of the groups proposed were: Ironman (red), Captain America (blue) or Thor (yellow).

The activity, following the definition of a game, is voluntary, the established rules must not be broken, it is carried out in a virtual framework (the story), and poses an artificial conflict (they must locate an object).

The dynamics of the game [47] were as follows.

- 1) The only *limitations* are temporal (2 h and a half).
- 2) Several *emotions* are felt (curiosity, creativity, competitiveness, frustration, happiness. . .) during the game.
- 3) The game *narrative* follows a guiding theme from a fantasy universe adapted for the students.
- 4) The students can see their *progress* in the subject.
- 5) Lasting *relationships* are established between the players, as will be seen in the activity evaluation (Section V).

The mechanics of the game [47] used were as follows.

- 1) *Cooperation*: They work in teams.
- 2) *Competition*: Only one team can win, although the best-classified teams can obtain extra points in the subject.
- 3) *Feedback*: The students know their mistakes in each task. After the activity, the exercises of the task were

corrected in class for all the students; however, some of them preferred to continue trying them on their own. This supports what will be seen in the results: the activity achieves its main goal, which is to improve the students' motivation and thus their learning process.

- 4) *Rewards*: The winning students obtain extra points in their marks and the object they were searching for during the game [Fig. 1(d)]. Furthermore, all the participants are rewarded with a chocolate medal of their color with the name of their team [Fig. 1(e)].
- 5) *Winning States*: Either when the number of teams to get extra points finish, or after two and a half hours.

The components of the game [47] used were as follows.

- 1) *Achievements*: Finding the desired object the first.
- 2) *Avatars*: Each team could choose the character with which they wished to participate.
- 3) *Levels*: Each task is more difficult than the previous.
- 4) *Teams*: The students worked together in groups.
- 5) *Quests*: They had to locate concrete elements (envelopes with the tasks) during the game.
- 6) *Social Graphic*: Each task has a different colored envelope that had to be kept on the team's table. So, how far through the game each team was is known.

The game was played in groups of 3 or 4 students. It was considered that a smaller size does not favor socialization. A *guided group formation* was used: the students were free to form their own groups, but had to choose academically heterogeneous partners, i.e., the members of a group could not be from the same degree or had to come from a different sphere of studies; while reinforcing socialization, the aim was to "mix" different levels of ability, favoring peer learning.

As it was a competitive activity, it was decided that the first three teams to finish would receive extra points: one for the winner (besides the trophy) and 0.5 for the second and third. It was satisfying to see that there were many participants who, after the winners had been declared and that therefore there were no extra points to play for, continued until they had finished all the tasks of the activity. At this point, it is important to note that, to be fair, the students who did not participate in the game could also get extra points in the subject by means of a test with the same problems as those included in the SH tasks.

To avoid frustration when they get stuck, an original element to the game was added: the use of a "joker." When a team was in a lower level than more than half the other teams, they could ask for a joker. This consisted of allowing one of the lecturers or assistants (students in higher courses) in the

activity to be with them for a maximum of 10 min to help with the problem, but without doing the program for them. It has been able to ascertain that this helped the less able students to avoid frustration, the *use of jokers* of some kind as part of the game is recommended.

#### IV. ACTIVITY EVALUATION: METHOD

##### A. Design

The study follows a *Cross-Sectional* design [48], [49], with qualitative and quantitative data collection.

The main aim with the activity is to change the student's attitude, but also for them to have a pleasant "user experience." So, according to Kirkpatrick's levels model [50], this part of the study focused on the students' *reaction*, i.e., level 1 of the model. Then, following what is usual in this kind of studies [23], a survey was used to collect the data. This survey was only done for the experimental group, i.e., the group who did the activity. In addition, there was also interest in the students' subjective perception about the learning of the topics included in the game tasks; so, a construct to measure this was also included in the survey.

However, there was also interest in objectively analyzing the possible influence of the game, more specifically the influence of the change in the attitude achieved with it, in the learning process. So, in this sense, the study was also performed at level 2 (*learning*) of the Kirkpatrick model. For this, and following what is also usual in [19] and [23], the data were collected by means of pre- and post-tests carried out at specific times during the course, with no control or intervention over the normal development of the course, i.e., the "variables" are observed, but without influencing them. This study follows a *Case-Control* design.

##### B. Participants

The subject in which the activity was developed has students of three different degrees: 1) computer science (CS); 2) statistics; and 3) a double degree in both. These students are randomly grouped in three different classes or forms. It is a subject of the first year of the Degree, so the participants have similar ages, in the range from 18 to 20. They are all from the same country and, although they may be from different regions, the sociocultural characteristics are similar, as well as their prior knowledge, since the great majority come from the same previous educational stage, except for a small percentage of students who were repeating the subject.

##### C. Sampling

An experimental protocol was followed [51], using experimental and control sets composed of students who did and did not do the activity, respectively. Following best practices [19], randomization has been used to assign participants to conditions. Two of the forms were randomly chosen to select, also randomly, students for the experimental set; the rest of the students formed the control set. The two sets consisted of 106 and 98 students, respectively. This is a high number when compared to the state of the art (Section II) and the minimum

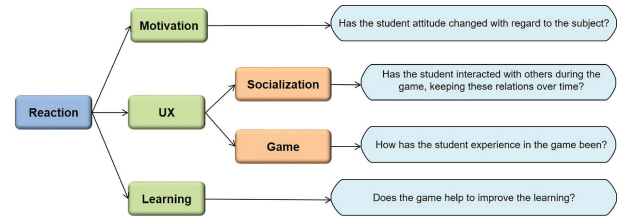


Fig. 2. Structure of the model used to evaluate the game.

participants per experimental set (condition) suggested in [19], allowing statistically significant results to be achieved.

##### D. Measures

The measures relating to the data to be collected during the study to achieve the evaluation objective were derived from the goals. Here, this definition was carried through by means of the hierarchical decomposition shown in Fig. 2. The evaluation was divided into four subcomponents or constructs. This model is an adaptation of the well-defined model for the evaluation of educational games (MEEGAs) [14], [52].

##### E. Hypothesis Formulation

The independent variable of the study was to do or not to do the activity, while the dependant ones were motivation, socialization, learning, and the student's experience. The following primary research hypotheses, aligned with the constructs defined in Fig. 2, were stated.

- 1) *Primary Hypothesis 1 (PH1)*: The serious game increases student motivation.
- 2) *Primary Hypothesis 2 (PH2)*: The serious game increases student socialization.
- 3) *Primary Hypothesis 3 (PH3)*: The students liked the activity.
- 4) *Primary Hypothesis 4 (PH4)*: The serious game boosts student learning.

The survey used was assessed by means of the following secondary hypotheses.

- 1) *Secondary Hypothesis 1 (SH1)*: The questions of each construct measure different aspects of the construct.
- 2) *Secondary Hypothesis 2 (SH2)*: The questions in the  $C_i$  construct are not correlated with those in the  $C_j$  construct, i.e., both constructs measure different aspects.

##### F. Data Collection

*Survey Implementation*: According to DeVellis [52], [53], the implementation of the measuring instrument must.

- 1) *Determine What to Measure*: This has already been done in previous paragraphs.
- 2) *Generate an Item Pool*: A set of 19 items aligned with the model defined in Fig. 2 were generated. Section V shows the questions for each construct.
- 3) *Determine the Format for Measure*: A Likert scale of five points was chosen, except for two of the items with multiple responses. So as not to condition the responses, the scale was masked: each question had five

responses that follow the Likert scale, but which were not numbered.

- 4) *Have the Initial Item Pool Reviewed by Experts?* The initial and following versions were iteratively reviewed and revised by experts in statistics, psychology, and educational innovation of the research group.
- 5) *Consider Inclusion of Validation Items:* No need was detected.
- 6) *Administer Items to a Development Sample and Evaluate the Items:* Since it was impossible to get a development sample that allows, for example, to calculate an interjudge agreement measure, this step was performed theoretically, matching the scale with well-defined ones [52], [54], [55].
- 7) *Optimize Scale Lengths:* The result of steps d) and f) is the final optimized version of the survey (Section V).

The survey was anonymous. In order not to be influenced by the “emotion of the moment,” “sufficient” time was allowed to elapse between the activity and the survey. Sufficient means enough time for the students to have become “settled” following the activity, but not so long as to forget what they had done or how they had felt. It was also wanted to check whether the relations forged during the game had passed the test of time. Defining such a delay exactly is not easy, but the survey was taken two weeks after the activity.

The survey fitted on a single sheet of paper and neither the items nor the possible responses were numbered.

*Pre/Post Tests Implementation:* The possible influence of the game in the learning process was evaluated in both the short and long term, using a pretest and two different “post-tests.” The pretest (PrT) was done about two weeks before the activity. To measure the short term influence, the students had another test about two weeks after the activity; which it is called the short term post-test (STPsT). These tests included the subject concepts up to that time. To measure the long term influence, the final mark of the subject was used. This is called the long term post-test (LTPsT). The final mark integrates all the tests taken during the course (including the final exam, which has the highest weight) and the laboratory work. As can be seen, it is not strictly a test, since it integrates diverse learning measures, but it does give a complete and better idea of the final competence of the student in the subject, which is what it is wanted to assess. All the tests were normal exams of the subject.

### G. Statistical Methods

1) *Survey Analysis:* For each construct in the survey, the correlation between the questions was calculated for the SH1 (Section IV-E) testing. The correlation between questions of different constructs was also analyzed (SH2). The *Spearman’s correlation coefficient* is used as the two variables in the correlation analysis have an ordinal scale. In this case, the null hypothesis used was  $H_0 : \rho_{\text{spearman}} = 0$ .

Each instructional activity goal was posed by means of a primary hypothesis. To validate these, the probability of achieving positive results (scale values 4 and 5) in each survey item was calculated, estimating their confidence interval at 95%. It is also calculated whether this probability of positive results was not achieved by chance, using a test for  $p$ ,

$H_0 : p = 0.5$ , i.e., the events happened randomly. This test is shown in the column “Significant?” of the results tables. In the case *non significant* results, the test power (TP) [56] was calculated.

2) *Pre-Post Test Analysis:* The results of both the STPsT and the LTPsT were compared with the corresponding ones in the pretest, which is the reference or baseline. The difference between the pre- and post-test marks ( $Dif = PostTest\_Mark - PreTest\_Mark$ ) and its relative value [ $RelDif = (Dif/PreTest\_Mark)$ ] was calculated for each student.

The mean and median of these values were calculated and compared to those of the experimental and control sets. The statistical significance of the difference was also measured here by means of the nonparametric *Mann–Whitney–Wilcoxon test*, *U-test*, since the distribution of the *Dif* and *RelDif* values does not follow a normal one. 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.

*Data Diagnostics:* The pre-/post-test analysis was done for all the students, grouping them by marks in the pretest (students that failed and passed it) to study the effect of this parameter and focusing on the “normal” student, i.e., eliminating the best and the worst who could condition the results; to do so, the best and worst 5% were eliminated, since this was considered a good compromise between removing the extremes, but maintaining an adequate experimental population.

## V. ACTIVITY EVALUATION: SURVEY RESULTS

The survey was answered by 101 of the 106 students in the experimental set (students who did the activity), since it was voluntary and not all the students were in class that day.

Before summarizing the results, it is important to note that the secondary hypotheses (Section IV-E) were validated.

- 1) *SH1:* The correlation analysis between questions showed very low values in all of the constructs.
- 2) *SH2:* Only the motivation and socialization constructs were considered, since these are the most related ones. All of the possible relations between the answers of the two constructs were studied. The correlation analysis showed not significant, in the majority of cases, or very low values. So, besides validating the survey, it can be seen that both are independent aspects: the use of serious games can improve motivation, but to improve socialization, the appropriate game dynamics to achieve it must be planned. This can also reinforce the initial approach, in which both are different goals to be achieved.

Fig. 3 shows the survey questions of motivation, socialization and user experience (game opinion) constructs. Here, for a better understanding, the survey questions are shown numbered and grouped by construct or subcategories. In the same way, the answers are also numbered from 1 (most negative opinion) to 5 (most positive) following a Likert-type scale, except questions 8 and 9 that are multiple response.

Following the hypothesis validation (Section IV-G1), the probability estimation of improvement (probability of answers

<p>1. Since doing the activity, I come to class...</p> <ol style="list-style-type: none"> <li>1. Now, I almost never attend class</li> <li>2. Less often than before</li> <li>3. With the same frequency</li> <li>4. More often than before</li> <li>5. Much more regularly than before</li> </ol>	<p>2. Since doing the activity, my attitude in class (I ask, answer, pay attention, ...) is...</p> <ol style="list-style-type: none"> <li>1. Much less active than before</li> <li>2. Less active than before</li> <li>3. The same as before</li> <li>4. A little more active than before</li> <li>5. Much more active</li> </ol>	<p>3. After doing the activity, I have done exercises by myself and I have studied programming...</p> <ol style="list-style-type: none"> <li>1. Much less active than before</li> <li>2. Less active than before</li> <li>3. The same as before</li> <li>4. A little more active than before</li> <li>5. Much more active</li> </ol>
(a)		
<p>13. I liked the activity...</p> <ol style="list-style-type: none"> <li>1. Not at all</li> <li>2. A little</li> <li>3. So so</li> <li>4. Quite a lot</li> <li>5. Very much</li> </ol>	<p>14. Doing the activity in a virtual environment, without leaving the laboratory</p> <ol style="list-style-type: none"> <li>1. Would have been much better</li> <li>2. Would have been better</li> <li>3. Would have been the same</li> <li>4. Would have been worse</li> <li>5. Would have been much worse</li> </ol>	<p>15. That the activity had a guiding narrative, in this case "the hunger games"...</p> <ol style="list-style-type: none"> <li>1. All that stuff confused me</li> <li>2. I could do without all that stuff.</li> <li>3. I don't mind.</li> <li>4. I preferred it like that. It was fun</li> <li>5. I liked very much. It seemed motivating</li> </ol>
(c)		
<p>4. During the game, I had a good time with other people</p> <ol style="list-style-type: none"> <li>1. I totally disagree</li> <li>2. I do not agree</li> <li>3. I neither agree nor disagree</li> <li>4. I agree</li> <li>5. I totally agree</li> </ol>	<p>5. During the game, I have had the opportunity to interact with the others</p> <ol style="list-style-type: none"> <li>1. I totally disagree</li> <li>2. I do not agree</li> <li>3. I neither agree nor disagree</li> <li>4. I agree</li> <li>5. I totally agree</li> </ol>	<p>6. The game encourages moments of cooperation and/or competition between players</p> <ol style="list-style-type: none"> <li>1. I totally disagree</li> <li>2. I do not agree</li> <li>3. I neither agree nor disagree</li> <li>4. I agree</li> <li>5. I totally agree</li> </ol>
<p>7. During the preparation and while playing the game...</p> <ol style="list-style-type: none"> <li>1. I have not interacted with anyone</li> <li>2. I have not met anyone new</li> <li>3. I have spoken to some people I had not met before</li> <li>4. I have met new people</li> <li>5. I have made new friends</li> </ol>	<p>8. [RM] Concerning my current relationship with the companions in my team, I can say that...</p> <ol style="list-style-type: none"> <li>1. We have become good friends.</li> <li>2. I have improved my relationship with all of them, but we haven't become close friends.</li> <li>3. With some, the relationship is better than before the activity.</li> <li>4. It is the same as before.</li> <li>5. With some, the relationship is much worse than before</li> </ol>	<p>9. [RM] Doing the activity has helped me to know much better...</p> <ol style="list-style-type: none"> <li>1. The Computer Engineering School</li> <li>2. The departments in the school</li> <li>3. The school staff</li> <li>4. The lecturers</li> <li>5. It has not helped me at all in this respect</li> </ol>
(b)		

Fig. 3. (a) Survey questions of motivation, (b) socialization, and (c) user experience (activity satisfaction) constructs.

TABLE I  
ESTIMATION OF THE PROBABILITY OF IMPROVEMENT (COLUMN  $P(imp)$ ) IN (A) MOTIVATION AND (B) SOCIALIZATION OR THAT "THE STUDENTS LIKE THE ACTIVITY" (COLUMN  $P(Like)$ ) IN (C). 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.*). WHEN THE ESTIMATION IS NOT STATISTICALLY SIGNIFICANT, THE TP VALUE IS SHOWN (A) MOTIVATION CONSTRUCT (B) SOCIALIZATION CONSTRUCT AND (C) GAME

(a)				(b)				(c)						
	N	$P(imp)$	C.I.	Significant?		N	$P(imp)$	C.I.	Significant?		N	$P(like)$	C.I.	Significant?
Q1	11	0.64	(0.36,0.92)	No ( $TP \leq 0.20$ )	Q4	93	0.94	(0.89,0.99)	Yes	Q13	80	0.97	(0.93, 1.00)	Yes
Q2	48	0.96	(0.91,1.00)	Yes	Q5	95	0.95	(0.91,1.00)	Yes	Q14	80	0.94	(0.88, 1.00)	Yes
Q3	82	0.99	(0.97,1.00)	Yes	Q6	100	0.99	(0.97,1.00)	Yes	Q15	91	0.98	(0.95, 1.00)	Yes
					Q7	54	0.91	(0.83,0.99)	Yes					

TABLE II  
PERCENTAGE OF STUDENTS THAT HAVE CHOSEN EACH ANSWER IN QUESTIONS 8 AND 9

	1	2	3	4	5
Q8	33%	45%	23%	22%	2%
Q9	51%	29%	9%	12%	26%

4 or 5) in motivation (PH1) and socialization (PH2) and satisfaction with the activity (PH3) measured in each question of the corresponding construct is shown in Table I. This table also shows the statistical significance of these improvements and "likes." The percentages of students that selected each response from questions 8 and 9 are shown in Table II.

In the survey, two questions were posed to relate the students' opinion about the activity to their opinion about the difficulty of the SH tasks and another two related to the students' gamer profile. The results in the first two showed that the students mostly believe that the exercises included are difficult or very difficult (70%) and that the time to do the activity is just right. With regard to the gamer profile, the answers showed a computer games player mainly (63% play computer games daily or every week), and that they rarely play noncomputer games (only 27% do so daily or every week).

## VI. ACTIVITY EVALUATION: PRE-/POST-TEST RESULTS

Here, the *learning construct* (Fig. 2) assessment is shown. The starting hypothesis is: "The serious game boosts students' learning," *primary hypothesis 4* (Section IV-E). This hypothesis was evaluated by objectively measuring the game's effect on their learning by means of pre- and post-tests (Section IV-F).

*Participant Flow*: The experimental set, (students who did the activity) had an initial size of 106. The control set (students

who did not do the activity) had an initial size of 98. From these students, 103 of the experimental set and 98 of the control one did the STPsT, and 91 and 86, respectively, the LTPsT (Section IV-F). Not all of the students that did the pre test did the post-tests, as some dropped out of the subject.

Table III shows the results of the pre-/post-tests study posed in Section IV-G2. The *U-test* for the small group of students who failed the pretest always shows the following warning: "cannot compute exact *p*-value." It is thought this is due to the small size of the test samples. So, the results in this case can be considered inconclusive and they are not considered in the final results discussion (Section VII).

## VII. DISCUSSION

In this section, the primary hypotheses in the light of the results shown in the previous sections are examined, so this section is organized following the construct model of Fig. 2.

### A. Motivation

The results shown in Table I(a) demonstrate that the use of the game does not seem to have any effect on class attendance [question 1, Fig. 3(a)]; this may be because it was already high before the activity. However, it has a positive effect on the students' attitude in class (question 2) and, even more, on their personal work (question 3), where most students pointed to improvements in doing exercises and personal study. These improvements are statistically significant, so PH1 is confirmed.

### B. Socialization

The results in Table I(b) show that the activity allowed the students not only to have fun with other people [question

TABLE III

RESULTS OF THE PRE-/POST-TESTS ANALYSIS. COLUMN *Students* SHOWS THE STUDENTS INCLUDED IN THE ANALYSIS: *All*, *No Pass Pre*, *Pass Pre*, and *Normal* MEAN THAT OF ALL THE STUDENTS IN THE EXPERIMENTAL AND CONTROL SETS, ONLY STUDENTS THAT FAIL THE PRETEST (PrT), ONLY STUDENTS THAT PASS THE PRETEST AND ELIMINATING THE 5% BEST AND WORST ARE USED, RESPECTIVELY. COLUMN *N* CONTAINS THE NUMBER OF STUDENTS IN EACH CASE. THE STATISTICAL SIGNIFICANCE (YES/NO,  $p$  – value < 0.05/ $p$  – value > 0.05) OF THE MEAN/MEDIAN DIFFERENCES (*ExpSet-CtrlSet* COLUMNS) IS SHOWN IN THE COLUMN *Significant?* AT A 95% CONFIDENCE LEVEL

Students	Experimental Set			Control Set			ExpSet - CtrlSet		U-test p-Value	Significant?
	N	Mean	Median	N	Mean	Median	Mean	Median		
<b>Short Term (Difference: <math>STPsT\_Mark - PrT\_Mark</math>)</b>										
All	103	1.07	1.17	91	0.69	0.67	0.38	0.50	0.059	No
No Pass Pre	13	1.90	1.91	20	1.36	0.89	0.54	1.02	0.29	No
Pass Pre	90	0.95	1.10	71	0.50	0.67	0.45	0.43	0.024	Yes
“Normal”	93	1.14	1.17	82	0.69	0.67	0.45	0.5	0.028	Yes
<b>Short Term (Relative Difference (in %): <math>\frac{STPsT\_Mark - PrT\_Mark}{PrT\_Mark} \cdot 100</math>)</b>										
All	103	33.6	39.0	91	20.4	20.8	13.2	18.2	0.013	Yes
No Pass Pre	13	27.2	31.8	20	13.1	13.7	14.1	18.1	0.37	No
Pass Pre	90	34.5	39.7	71	21.6	20.3	12.9	19.4	0.024	Yes
“Normal”	93	34.7	39.0	82	23.4	22.1	11.2	16.8	0.004	Yes
<b>Long Term (Difference: <math>LTPsT\_Mark - PrT\_Mark</math>)</b>										
All	98	-1.54	-1.0	86	-1.94	-1.33	0.41	0.33	0.18	No
No Pass Pre	12	-1.72	-1.77	16	-1.16	-1.56	-0.56	-0.21	0.6	No
Pass Pre	86	-1.51	-0.80	70	-2.12	-1.33	0.61	0.53	0.08	No
“Normal”	88	-1.51	-1.03	78	-2.15	-1.50	0.64	0.48	0.049	Yes
<b>Long Term (Relative Difference (in %): <math>\frac{LTPsT\_Mark - PrT\_Mark}{PrT\_Mark} \cdot 100</math>)</b>										
All	98	-24.0	-14.3	86	-32.5	-24.4	8.5	10.1	0.15	No
No Pass Pre	12	-54.1	-65.6	16	-38.4	-38.1	-15.7	-27.5	0.5	No
Pass Pre	86	-19.8	-11.6	70	-31.2	-18.6	11.4	7.0	0.049	Yes
“Normal”	88	-27.2	-14.4	78	-36.9	-27.3	9.7	12.9	0.039	Yes

4, Fig. 3(b)], but also to interact (questions 5 and 6) and meet new people (question 7). In addition, a high percentage of these relationships have been maintained over time (question 8): many more students selected answers 1, 2, and 3 than those who selected answers 4 and 5. Concerning the relation with the educational environment (question 9), a high percentage of students stated that the activity helped them to know the School better. The results are statistically significant, which has allowed to statistically prove the improvements in the socialization hypothesis (PH2).

This is very interesting, since socialization, which is important in all stages of education, is especially so for the first course university students, as most did not know each other previously and for whom the start of this new period of their formative life is crucial. Socializing allows the students to feel comfortable as part of the group. It is also very important to acquire confidence in the educational environment in which they will develop over the following few years, in both physical and emotional interpersonal relationships.

### C. Game

From the results, it can be concluded that the primary hypothesis related to the game (PH3) has been statistically validated [Table I(c)]. The students like the activity (question 13) and they prefer the non virtual nature of the game (question 14); it is interesting to note that only 5% selected the options 1 and 2 in question 14. The use of a narrative (and the elements that accompany it) to create an atmosphere to engage the players supposes extra time and cost in the preparation. It is therefore important to see if it is worth the effort. The results confirm that the answer is clearly yes (question 15).

These positive opinions are even more interesting considering the students think that the game tasks are difficult and that their gamer profile is mainly that of a video game player (Section V).

### D. Learning

As the previous improvements in the students' attitude and their high satisfaction with the activity is important, its usefulness would not be complete without quantitative evidence to support its effectiveness in the students' learning process. This has been shown in the pre-/post-test results (Section VI).

The first important conclusion from Table III is that the results in the post-test concerning the pretest were always better (except for students who failed the pretest in the long term study) for the students in the experimental set than for those in the control set.

The second important conclusion is that the effect of the activity on the learning in the short term appears to be positive. The results of the experimental set are always better, and this difference is statistically significant, except for the *difference* for *All* the students, but with a very low- $p$ -value and close to the 0.05 limit, and for those who did not pass the pretest, whose results are not conclusive, as pointed out previously.

As for the long term learning study, the first aspect to note is that the final marks are worse, in general, than the pretest ones. This is usual, since the subject's final mark contains all of the tests and works done along the school year, including the final exam, which is the test with the highest weight in the final mark and all the concepts of the subject are included. Taking this into account, Table III shows that better, statistically significant, results were achieved for the experimental set when the extremes (“best and worst students”) are eliminated. This is also observed for the students who passed the pretest when the relative difference is calculated. Both cases supposed the great majority of the students, so it can be concluded that the activity also appears to have a positive effect on learning in the long term for most of the students.

All of the above has allowed to conclude that the PH4 is also confirmed. This shows that the present proposal, and



similar activities, focused on motivation and improving the class atmosphere (socialization), can positively influence the students' learning.

## VIII. THREATS TO VALIDITY

The potential threats which may affect the validity of the study results can be divided into construct, internal, external, and conclusion validity [57]. In the following, these issues are addressed, describing how these threats were approached in order to avoid them in the work.

### A. Construct Validity

The main threats to ensure that what is measured is what is intended to be measured come from the design of the instruments. In this sense, the survey was designed and validated by experts following well-defined and validated models [52], [54], [58] adapted to the differential characteristics of the proposal. The survey was also assessed, confirming that different questions within a construct measure different aspects, and that questions of different constructs are not correlated. All the outcomes were evaluated. As for the objective measure of the learning, the pre- and post-tests were not created ad hoc, with the risk of introducing subjectivity; normal exams, independent of the activity and created to assess the student's competence in the subject, were used instead.

### B. Internal Validity

The main threats here are due to the data collection conditions and the limitations inherent to the cross-sectional design followed [59]. It is important that the data should be as representative as possible with regard to the population under study and collected as objectively as possible to avoid bias.

To achieve this, an experimental protocol was first followed with experimental and control sets, created using random sampling. Both sets have the same characteristics, since they belong to the same degree courses with exactly the same learning and assessment activities and study materials. Although the subject has two different teachers, they have the same experience and knowledge. The activity evaluation instruments (survey and pre-/post-tests) were the same for all the students, these being representative of the study population.

Focusing on the implementation, special attention was paid to guaranteeing the validity of the acquisition measures, so as to be as objective as possible. As for the survey, most importantly, it was anonymous. The scale of the questions and their division into constructs was masked from the students. So as not to be influenced by the "emotion of the moment," the survey was taken two weeks after the activity. The students were not trained for either the survey or the tests and their responses were not conditioned by the researchers (*Rosenthal effect*); moreover, most of the researchers were not present during these assessment activities, since they were not the subject teacher, thus avoiding introducing any bias.

So that the *extra points* the participants in the activity could obtain did not affect the voluntary participation and, therefore, the study results; prior to signing up for the activity, the

students were informed that they could achieve, if they wanted, the same *extra points*, but, with a test that included the same problems as those included in the game tasks.

Finally, it is important to note that the researcher who performed the activity and collected the data was different from those who analyzed them, and did not intervene in the analysis to avoid bias. Moreover, the researchers who performed the data analysis were not teachers of the subject, nor did they know the students.

### C. External Validity

The results cannot be generalized yet, as the evaluation was carried out in our university only. However, given the nature of the proposal and the profile of the participants, it is believed that the results can be extrapolated to similar sociocultural and educational conditions, but also to different ones, since the literature shows the successful use of SH in completely different cultural and educational environments [33], [34], [41], [43], [60], [61].

Students from three different degrees participated in the activity and, in addition, students of higher courses who asked to carry out the activity without opting for a reward. This would seem to show that it is not only an interesting and attractive activity for CS studies and for students in the first university years, though this is left for future experimentation.

### D. Conclusion Validity

Conclusion validity concerns those aspects that might affect the ability to draw a correct conclusion from the statistical analysis of the data. The main threat here came from the data collection (sampling, size, representativeness, etc.), the use of appropriate statistical tests and the reliability of the measurement.

Most of the issues concerning the validity of the data collection have been stated in the previous sections. Here, it is only wished to add that the sizes of the experimental and control sets, besides achieving statistically significant results, are higher than those in the majority of the related bibliography and with the minimum of participants suggested in [19].

The study is based on the assessment of hypotheses. This evaluation was performed by means of measurements with proven efficacy in acquiring students' reactions (survey). They were also designed following well-defined and validated models; using instruments with proven efficacy in measuring student competence, as well as the use of pre- and post-tests. The results were statistically analyzed using suitable methods for each measure. In general, these results are statistically significant, but the cases in which nonsignificant ones were achieved, this has been explicitly mentioned; the power of the statistical test being calculated, whenever possible, in these cases.

## IX. LIMITATION

The subject has two different lecturers. Although this is not desirable for a more uniform study, it was impossible to avoid due to teaching distribution issues external to the activity.

However, as pointed out, the effect of this limitation was reduced because both lecturers have the same preparation and experience in the subject. Furthermore, the learning materials and evaluation tests, which are the same for all the students in the subject, were created jointly by both lecturers.

The experimental and control sets were made up of students studying the same subject; so they may have influenced each other. Although it is impossible to eliminate this, it has been tried to limit this influence in the study by choosing the students for the experimental set from two of the forms, leaving the third for the control one only. Furthermore, the tests used in the pre-/post-test study are individual tasks.

Another limitation is that all the students belong to the same country, with a similar socioeconomic and cultural condition. However, they are representative of the study population and comparable with students of other sites, but with similar cultural and educational conditions.

The temporal restrictions are another external limitation to the research. Both the activity and its assessment have been adapted to affect the normal development of the subject as little as possible, in order not to overload the student.

## X. CONCLUSION

In this article, a SH for a nondigital serious game in higher education has been described and systematically evaluated.

An ample review of the state of the art has been performed focused on the use of serious games in higher education, in general, and on the use of SH in particular. From this review, it can be concluded that the use of serious games in higher education is a growing trend, these being mostly videogames, unlike the one proposed here. Several experiences about the use of SH in the University have been also found, but none with the approach shown here and with limitations in the assessment of outcomes.

An empirical evaluation of the activity has been presented, based on a survey to obtain the opinion of the students in the experimental set concerning the desired goals, as well as the activity itself; the evaluation also included pre- and post-tests, with experimental and control sets. All of the results have been statistically analyzed.

These results suggest that the activity reached the goals posed.

- 1) To increase the student's engagement with the subject (PH1). The students showed an improvement in their motivation, being more proactive and participatory.
- 2) To encourage socialization (PH2). The students indicated that most of them established new relations during the activity, these being maintained over time.
- 3) To boost learning (PH4). The activity allowed the student to know their educative environment (school, teachers, etc.) better, making them feel more comfortable; in addition, the game allowed them to establish new relations with their classmates, supporting peer learning and increasing motivation. All these factors mean that the students become more involved in their studies, which seems to favorably influence their learning.

With these three parameters approached through the instructional activity, the aim was to deal with the emotions and behavior that are the enemies of learning: fear, isolation, boredom, anxiety, impotence, indolence, etc. Motivated students, within a peer group in which everyone learns from everyone, integrated in the educational environment and confident of their capacities and abilities, have more possibilities of growing, both academically and personally, as can be concluded from the results of the study.

Furthermore, the students showed that they had a nice and enjoyable experience (PH3). The majority of the students liked the activity, valuing very positively the fact that this was non digital.

The very positive results support the teachers' observations, encouraging us to continue in this line of work, since it favors the integration and engagement of the students with the subject.

## REFERENCES

- [1] R. Branch, *Instructional Design: The ADDIE Approach*. New York, NY, USA: Springer, 2009.
- [2] S. Kurt, "Addie model: Instructional design," 2017. [Online]. Available: <https://educationaltechnology.net/the-addie-model-instructional-design/>
- [3] A. Min, H. Min, and S. Kim, "Effectiveness of serious games in nurse education: A systematic review," *Nurse Educ. Today*, vol. 108, Jan. 2022, Art. no. 105178.
- [4] J. Zeng, S. Parks, and J. Shang, "To learn scientifically, effectively, and enjoyably: A review of educational games," *Human Behav. Emerg. Technol.*, vol. 2, no. 2, pp. 186–195, 2020.
- [5] S. Subhash and E. A. Cudney, "Gamified learning in higher education: A systematic review of the literature," *Comput. Hum. Behav.*, vol. 87, pp. 192–206, Oct. 2018.
- [6] C. A. Bodnar, D. Anastasio, J. A. Enszer, and D. D. Burkey, "Engineers at play: Games as teaching tools for undergraduate engineering students," *J. Eng. Educ.*, vol. 105, no. 1, pp. 147–200, 2016.
- [7] M. Qian and K. R. Clark, "Game-based learning and 21st century skills: A review of recent research," *Comput. Hum. Behav.*, vol. 63, pp. 50–58, Oct. 2016.
- [8] E. Naul and M. Liu, "Why story matters: A review of narrative in serious games," *J. Educ. Comput. Res.*, vol. 58, no. 3, pp. 687–707, 2020.
- [9] J. D. Gomez-Escalonilla-Torrijos, "Motor motivation. A neuroeducational strategy to improve the active participation of students in their learning and generate subjective perception of happiness," Ph.D. dissertation, Faculty Educ., Universidad Camilo José Cela, Madrid, Spain, 2017.
- [10] A. M. Soledispa-Rivera, E. J. San-Andres-Soledispa, and R. A. Soledispa-Pin, "Motivation and its influence on the academic performance of students in basic higher education," *Sinapsis*, vol. 3, no. 18, Dec. 2020.
- [11] H. Almas, F. Pinkow, and F. Gæver, "Reimagining how to understand learning game experiences: A qualitative and exploratory case study," *Int. Res. J. Eng., IT Sci. Res.*, vol. 10, no. 14, pp. 1–26, 2023.
- [12] 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," *Amer. Psychol.*, vol. 73, no. 1, pp. 3–25, 2018.
- [13] P. E. Battistella and C. G. Von Wangenheim, "Games for teaching computing in higher education: A systematic review," *IEEE Technol. Eng. Educ.*, vol. 9, no. 1, pp. 8–30, Mar. 2016.
- [14] G. Petri and C. G. Von Wangenheim, "How games for computing education are evaluated? A systematic literature review," *Comput. Educ.*, vol. 107, pp. 68–90, Apr. 2017.
- [15] R. K. Stark, E. Opuda, J. McElfresh, and K. Kauffroath, "Scavenging for evidence: A systematic review of scavenger hunts in academic libraries," *J. Acad. Librariansh.*, vol. 47, no. 3, May 2021, Art. no. 102345.
- [16] L. H. Taraldsen, F. O. Haara, M. S. Lysne, P. R. Jensen, and E. S. Jenssen, "A review on use of escape rooms in education—Touching the void," *Educ. Inquiry*, vol. 13, no. 22, pp. 169–184, 2020.

- [17] A. Asgari, "Impact assessment as a new horizon for assessing the game's intervention: A framework for the impact assessment of educational games," in *Proc. ISGS*, 2021, pp. 18–26.
- [18] D. Vlachopoulos and A. Makri, "The effect of games and simulations on higher education: A systematic literature review," *Int. J. Educ. Technol. Higher Educ.*, vol. 14, no. 1, p. 22, 2017.
- [19] A. All, E. P. N. Castellar, and J. V. Looy, "Assessing the effectiveness of digital game-based learning: Best practices," *Comput. Educ.*, vols. 92–93, pp. 90–103, Jan./Feb. 2016.
- [20] T. H. Laine and R. S. N. Lindberg, "Designing engaging games for education: A systematic literature review on game motivators and design principles," *IEEE Trans. Learn. Technol.*, vol. 13, no. 4, pp. 804–821, Oct.–Dec. 2020.
- [21] M. Dorriv, "Incorporation of serious games into higher education: A survey," in *Proc. ISGS*, 2021, pp. 86–90.
- [22] S. De. Freitas, "Are games effective learning tools? A review of educational games," *Educ. Technol. Soc.*, vol. 21, no. 2, pp. 74–84, 2018.
- [23] E. A. Boyle et al., "An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games," *Comput. Educ.*, vol. 94, pp. 178–192, Mar. 2016.
- [24] Z. K. Westenhaver, R. E. Africa, R. E. Zimmerer, and B. J. McKinnon, "Gamification in otolaryngology: A narrative review," *Laryngoscope Investig. Otolaryngol.*, vol. 7, no. 1, pp. 291–298, 2022.
- [25] H. Ning, H. Wang, W. Wang, X. Ye, J. Ding, and P. Backlund, "A review on serious games in e-learning," 2022, *arXiv:2201.06917*.
- [26] C. Wang and L. Huang, "A systematic review of serious games for collaborative learning: Theoretical framework, game mechanic and efficiency assessment," *Int. J. Emerg. Technol. Learn. (iJET)*, vol. 16, no. 6, pp. 88–105, Mar. 2021.
- [27] P. Sajjadi, A. Ewais, and O. De Troyer, "Individualization in serious games: A systematic review of the literature on the aspects of the players to adapt to," *Entertain. Comput.*, vol. 41, Mar. 2022, Art. no. 100468.
- [28] M. Arevalo, V. Romero, C. Sanchez-Zhunio, D. Prado-Cabrera, and P. Cedillo, "Towards the evaluation of the usability and portability of serious games," in *Proc. IEEE 9th Int. Conf. Ser. Games Appl. Health(SeGAH)*, 2021, pp. 1–8.
- [29] A. D. Ronan, "A scavenger hunt activity to welcome first-year students to the civil engineering department," in *Proc. ASEE Annu. Conf. Expo. (ASEE)*, Tampa, Florida, 2019, pp. 1–13.
- [30] S. Eric Nybo et al., "Design of a large-scale escape room for first-year pharmacy student orientation," *Curr. Pharm. Teach. Learn.*, vol. 12, no. 11, pp. 1340–1347, Nov. 2020.
- [31] M. Fakhour, A. Azough, F.-Z. Kaghat, and M. Meknassi, "A cultural scavenger hunt serious game based on audio augmented reality," in *Proc. Int. Conf. Adv. Intell. Syst. Sustain. Develop.*, 2020, pp. 1–8.
- [32] A. Hutzler, R. Wagner, J. Pirker, and C. Gütl, "Mythhunter: Gamification in an educational location-based scavenger hunt," in *Proc. 3rd Int. Conf. iLRN*, 2017, pp. 155–169.
- [33] J. Nessler, E. Schaper, and A. Tipold, "Proof of concept: Game-based mobile learning—the first experience with the app actionbound as case-based geocaching in education of veterinary neurology," *Front. Vet. Sci.*, vol. 8, Dec. 2021, Art. no. 753903.
- [34] D. Alajaji and A. A. Alshwiah, "Effect of combining gamification and a scavenger hunt on pre-service teachers' perceptions and achievement," *J. Inf. Technol. Educ. Res.*, vol. 20, pp. 283–308, Jun. 2021.
- [35] L. Li, "Augmented reality facilitated scavenger hunt for mobile learning," *J. Scholarsh. Teach. Learn.*, vol. 20, no. 2, pp. 113–121, 2020.
- [36] E. Katsadourous, D. Kogias, L. Toumanidis, C. Chatzigeorgiou, and C. Z. Patrikakis, "Teaching network security through a scavenger hunt game," in *Proc. IEEE Glob. Eng. Educ. Conf. (EDUCON)*, 2017, pp. 1802–1805.
- [37] C. Löw and L. Moshuber, "Grätzelbot—Gamifying onboarding to support community-building among university freshmen," in *Proc. 11th Nordic Conf. Hum.-Comput. Interact. Shap. Exp., Shap. Soc.*, New York, NY, USA, 2020, pp. 1–3.
- [38] P. Williams, "Using escape room-like puzzles to teach undergraduate students effective and efficient group process skills," in *Proc. IEEE Integr. STEM Educ. Conf. (ISEC)*, 2018, pp. 254–257.
- [39] M. Dyrud, "The scavenger hunt: A team building exercise," in *Proc. ASEE Annu. Conf. Expo.*, 2008, pp. 13–1263.
- [40] M. Wolf, F. Wehking, H. Sobke, and J. Londong, "Virtualization of virtual field trips: A case study from higher education in environmental engineering," in *Proc. DELBA*, 2020, pp. 1–11.
- [41] M.-C. Corujo-Vélez, M.-T. G. Del Castillo, and A.-E. Merla-González, "Construtivist and collaborative methodology mediated by ICT in higher education using webquest," *Pixel-Bit*, vol. 57, pp. 7–57, Jan. 2020. [Online]. Available: <https://recyt.fecyt.es/index.php/pixel/article/view/70422>
- [42] E. R. Bekeš, "Creating multisensory learning experiences that go beyond the limitations of traditional media," in *Proc. 44th Int. Conv. Inf. Commun. Electron. Technol. (MIPRO)*, 2021, pp. 587–590.
- [43] S. Kanneiger, "Effects of an augmented reality library orientation on anxiety and self-efficacy: An exploratory study," *Coll. Res. Libr.*, vol. 82, no. 3, pp. 352–369, 2021.
- [44] A. Manzoor, M. Samarin, D. Mason, and M. Ylianttila, "Scavenger hunt: Utilization of blockchain and IoT for a location-based game," *IEEE Access*, vol. 8, pp. 204863–204879, 2020.
- [45] S. Laato, T. Pietarinen, S. Rauti, M. Paloheimo, N. Inaba, and E. Sutinen, "A review of location-based games: Do they all support exercise, social interaction and cartographical training?" in *Proc. 11th Int. Conf. Comput. Support. Educ. (CSEDEU)*, 2019, pp. 616–627.
- [46] G. R. Greher and S. H. Marshall, "The app scavenger hunt: Helping future music educators embrace change," in *Creative Music Making at Your Fingertips: A Mobile Technology Guide for Music Educators*. Oxford, U.K: Oxford Scholarsh. Press, 2021, pp. 56–66.
- [47] K. Werbach and D. Hunter, *The Gamification Toolkit: Dynamics, Mechanics, and Components for the Win*. Philadelphia, PA, USA: Warton Digit. Press, 2015.
- [48] I. Anton-Solanas 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 Educ. Today*, vol. 118, Nov. 2022, Art. no. 105527. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0260691722002635>
- [49] K. Bovermann and T. J. Bastiaens, "Towards a motivational design? Connecting gamification user types and online learning activities," *Res. Pract. Technol. Enhanc. Learn.*, vol. 15, no. 1, pp. 1–18, Jan. 2020.
- [50] D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating Training Programs*, 3rd ed. San Francisco, CA, USA: Berrett-Koehler Publ., 2006.
- [51] G. Petri and C. Von Wangenheim, "How to evaluate educational games: A systematic literature review," *J. Univers. Comput. Sci.*, vol. 22, no. 7, pp. 992–1021, 2016.
- [52] R. Savi, C. G. Von Wangenheim, and A. F. Borgatto, "A model for the evaluation of educational games for teaching software engineering," in *Proc. 25th Braz. Symp. Softw. Eng.*, 2011, pp. 194–203.
- [53] R. F. DeVellis, *Scale Development: Theory and Applications*, 2nd ed. Thousand Oaks, CA, USA: Sage Publ., 2003.
- [54] F.-L. Fu, R.-C. Su, and S.-C. Yu, "EGameFlow: A scale to measure learners' enjoyment of e-learning games," *Comput. Educ.*, vol. 52, no. 1, pp. 101–112, 2009.
- [55] B. K. S. Utvar and G. Haugan, "The academic motivation scale: Dimensionality, reliability, and construct validity among vocational students," *Nordic J. Vocat. Educ. Train.*, vol. 6, no. 2, pp. 17–45, Nov. 2016. [Online]. Available: <https://njetvet.ep.liu.se/article/view/3218>
- [56] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Stratford, ON, Canada: LEA Publ., 1988.
- [57] C. Wohlin, P. Runeson, M. Host, M. C. Ohlsson, B. Regnell, and A. Wesslen, *Experimentation in Software Engineering*, 1st ed. Berlin, Germany: Springer, 2012.
- [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," *Comput. Hum. Behav.*, vol. 71, pp. 525–534, Jun. 2017.
- [59] M. S. Setia, "Methodology series module 3: Cross-sectional studies," *Indian J. Dermatol.*, vol. 61, no. 3, pp. 261–264, 2016.
- [60] P. Nanayakkara, T. Rathnaweera, S. Weerasingha, Y. Pieris, and R. Rajapakse, "Interactive game application for garbage disposal using image processing and location based services," in *Proc. IEEE ICIAfS*, 2018, pp. 1–5.
- [61] B. Mei and S. Yang, "Nurturing environmental education at the tertiary education level in china: Can mobile augmented reality and gamification help?" *Sustainability*, vol. 11, no. 16, p. 4292, 2019.

**Alma Pisabarro-Marron** received the B.Sc. degree in computer science from the University of Valladolid, Valladolid, Spain, in 1996.

She is currently an Associate Professor with the University of Valladolid. She has led several research projects and is currently a Co-Coordinator of the GREIDI Teaching Innovation Group (<https://greidi.infor.uva.es/>). Her fundamental research line is teaching innovation, mainly gamification, with several publications in conferences and journals.

**Carlos Vivaracho-Pascual** received the Bachelor of Science degree in physics and the Ph.D. degree from the University of Valladolid, Valladolid, Spain, in 1989 and 2004, respectively.

He is currently a Lecturer with the University of Valladolid. He has published numerous papers in relevant congresses and journals and some book chapters (<https://www2.infor.uva.es/cevp/publicaciones.php>). His research interests are both biometric and educational innovation.

**Silvia Arias-Herguedas** received the B.Sc. degree in computer science from the University of Valladolid, Valladolid, Spain, in 2018, and the master's degree in big data and business intelligence from the University of Leon, León, Spain, in 2021.

She was an Assistant Professor with the University of Valladolid. She is currently a DevOps Engineer for an international company. She is interested in cloud infrastructure, cybersecurity, learning analytics, and gamification.

**Esperanza Manso-Martinez** received the Bachelor of Science degree in mathematics and the Ph.D. degree from Valladolid University, Valladolid, Spain, in 1977 and 2009, respectively.

She was a Lecturer with the Department of Computer Language and Systems, Valladolid University. She is interested in software maintenance, reuse, and experimentation. Her works have been accepted in international journals and congresses.