International Journal of Comparative Education and Development

A multistage linear path design for educational escape rooms

Pedro Juan Roig, Salvador Alcaraz, Katja Gilly and Cristina Bernad
Department of Computer Engineering, Miguel Hernández University, Elche, Spain, and
Carlos Juiz
Department of Mathematics and Computer Science, University of the Balearic Islands, Palma de Mallorca, Spain

Received 1 September 2025
Revised 28 January 2026
16 March 2026
Accepted 17 March 2026

Abstract

Purpose – Active learning methodologies allow students to bring a more active role in their education, where active evaluation tools may play an important part. In this paper, a multistage linear path design for educational escape rooms is described, where a particular level must be cleared before moving on to the following one. This way, each level may be dedicated to a different didactic unit within the curriculum.

Design/methodology/approach – The scheme of the design proposed is first described in an informal way, focused on the Spanish grading system. Afterwards, the basics of the design are abstracted away in order to describe it in a formal way, aimed at a generic grading system. The escape room is used as a group-based active evaluation tool.

Findings – The outcome obtained with this scheme provided an 11% increase in academic performance and a 17% increase in success rate with respect to the previous academic year, where evaluation was done through a traditional written exam. Moreover, the level of engagement achieved was high, according to the ISA engagement scale. Additionally, a *t*-test was performed with the distribution of scores obtained in both courses, showing that the results were statistically significant. However, the sample size was insufficient; hence, further research should be carried out with a larger group.

Originality/value – The novelty of this approach is in the format of the escape room, whose design is a multistage linear path.

Keywords Active learning, Escape room, Engagement level, Gamification, Innovative education

Paper type Research article

1. Introduction

Education 4.0 focuses on providing learners with knowledge, skills, attitudes and competences required to be able to adapt to a rapidly evolving digital society, where technological changes take place faster and faster (Mukul and Büyüközkan, 2023). Hence, students need to develop key concepts to keep pace with the technology landscape, such as self-learning, problem-solving, leadership or critical thinking (Matsumoto-Royo *et al.*, 2021). All those features are integrated into the learning to learn paradigm, where the key point is the development of learning strategies so as to optimize the adaptation time to new technological environments, which currently happens at an ever-increasing pace (Lansdell and Kording, 2019). Therefore, self-learning environments must be presented to the pupils in order for them to get used to the ever-growing changes led by new techno paradigms (Pacheco-Velázquez *et al.*, 2024).

Considering the self-learning requirements, active learning is one of the best ways to go in order to implement Education 4.0, as it accounts for a paradigm shift related to how the

Funding: This research received no external funding.

Conflict of interest statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

International Journal of Comparative Education and Development
Emerald Publishing Limited
e-ISSN: 2309-4907
p-ISSN: 2396-7404
DOI 10.1108/IJCED-09-2025-0121

learning process takes place (Gueye and Expósito, 2023). Actually, in active learning methods, students take an active role when it comes to their education, whilst teachers take a more passive role in order to dynamize the learning process driven by the students (González-Pérez and Ramírez-Montoya, 2022).

One of the key points in active learning methodologies is gamification, which can be defined as using game design elements for non-game contexts (Deterding *et al.*, 2011). In this sense, Gee influences instructional design by advocating for game-based elements that pair learning goals with engagement (Gee, 2003). Some relevant aspects include contextualized learning, on-demand information, self-environment for failure and game mechanics as pedagogy (Gee, 2005).

Gamification can be implemented in different ways, such as through the use of edu-breakout rooms as strategies to enhance engagement and learning (Edu-Agustín *et al.*, 2025), the use of gamified environments in which websites serve as the primary technology for delivering gameful experiences (Oliveira and Hamari, 2025), or the deployment of gameful assessment and progression mechanics (Bipp *et al.*, 2024). In general, gamification in education is considered a mechanism that tends to improve both motivation and learning outcomes (Rivera and Garden, 2021).

The effectiveness of active learning methodologies is grounded in well-established theoretical constructs such as motivation, autonomy and collaborative learning. From the perspective of self-determination theory, intrinsic motivation and students' perceived autonomy have been shown to play a central role in enhancing engagement and academic performance (Ryan and Deci, 2000). Likewise, empirical evidence indicates that active learning strategies lead to significant gains in learning and reductions in failure rates compared to traditional approaches, particularly when they promote active participation and self-regulation (Michael, 2006).

On the other hand, collaborative learning has been widely associated with cognitive and socioemotional improvements, as it facilitates positive interdependence and the shared construction of knowledge (Johnson and Johnson, 2018). Taken together, a more explicit integration of these constructs would help strengthen the theoretical framework that explains the effectiveness of active learning methodologies in educational contexts.

In this context, the research problem in this paper is the need to implement a more self-driven approach in the learning process in a course dedicated to learn the basics of networking, whereas the research question proposed is whether it is possible to deploy an active learning method in this kind of course. The goal is to achieve better academic results driven by a boost in self-learning, as well as to obtain a high level of engagement among learners (Peláez-Sánchez *et al.*, 2024).

There are different types of active learning methods, such as project-based learning, team-based learning or serious gaming, where academic results are often improved compared to traditional learning (Patiño *et al.*, 2023). However, the use of educational escape rooms provides both the self-directed learning component and its gamification component, which adds a motivational boost to the enhanced results provided by other active learning methods (Stieha *et al.*, 2024). Hence, the use of educational escape rooms provides an ideal approach for reframing the research question as to whether it is possible to deploy an educational escape room in a particular course devoted to networking essentials in order to achieve both an improvement in academic performance and an enhancement in the level of engagement among learners (Blanco *et al.*, 2023).

The benefits of using educational escape rooms have been reported across various areas and levels in the education field. Focusing on research articles published in 2025, Ugo *et al.* reported improved academic performance when applying this methodology to programming courses in higher education (Ugo *et al.*, 2025), whereas Liu and Liu found higher levels of student engagement in Business Administration (Liu and Liu, 2025). Similarly, Sarin *et al.* highlighted their pedagogical impact on the development of soft skills in Healthcare education (Sarin *et al.*, 2025), whilst Yu *et al.* underlined their potential to enhance technical skills in

Nursing education (Yu *et al.*, 2026). However, these benefits are not constrained to higher education, as Grande de Prado *et al*. described their role in developing competences in primary education (Grande de Prado *et al.*, 2025), while Giner-Baixauli *et al*. emphasized their contribution to knowledge acquisition and motivation in secondary education (Giner-Baixauli *et al.*, 2025).

In this paper, a multistage linear path structure is presented for educational escape rooms, with the particularity that each stage must be first completed before jumping into the next stage, thus accounting for an independent collection of linear paths. This way, each of those stages could be dedicated to the evaluation of a given didactic unit, which allows the use of this type of escape room to implement an active evaluation tool where a range of didactic units could be assessed altogether.

The organization of the rest of the paper is the following: Section 2 presents the method, Section 3 describes the results obtained, Section 4 presents the discussion about those results, and Section 5 presents the final conclusions.

2. Method

There are different options when designing an escape room, either educational or recreational. The simplest option would be "linear path," where all puzzles in the path must be traversed sequentially until the final puzzle, also known as the metapuzzle, is reached. This way, there is only one possible path to be followed by all players with that layout. Another option would be "open path," where there are alternative lines to go through the escape room until the metapuzzle is reached, such that there are alternative paths for all players (Nicholson, 2015).

A combination of both architectures would lead to a "path-based" layout, where multiple paths could be sequentially traversed. Furthermore, any other hybrid model is also possible, where different paths might cross at certain points in order to take different routes so as to reach the metapuzzle and overcome it, which stands for the completion of the game (Soares-Collado and Sánchez-Hernández, 2020).

However, in this paper, we propose a "multistage linear" path, which could be seen as a combination of linear paths to be traversed in a sequential order. This way, each stage is composed of a linear path, where it is necessary to clear the current path in order to start with the next path, such that the game is completed when all levels have been cleared.

It should be noted that in multistage linear structures, the predefined order and explicit milestones often reduce uncertainty, which can help lower anxiety by providing clear expectations and a sense of progression. This way, students know what the next step is and what is required to advance. On the other hand, the open path models offer greater autonomy and flexibility but may increase cognitive load and anxiety for some learners, particularly those who prefer structured guidance, because they must decide which path to follow and how to manage their progress. Therefore, while open-path approaches can foster creativity and self-regulation, multistage linear designs may provide a more supportive framework for students who benefit from clear structure and incremental goals.

In the following subsections, the structure proposed is going to be described as an informal description and as a formal representation.

2.1 Informal representation

To start with, each escape room is structured as a series of sequential stages where each of those is composed of a bunch of sequential states. In other words, such a structure could be seen as a sort of bidimensional array (Rivest and Ebouele, 2020), where each row, also known as a level, must be traversed horizontally, going from left to right through the different columns therein, also known as states. This way, the overall target of this layout is to go through all levels in a sequential manner as soon as possible, where each level must be completed. Hence, no vertical movements are allowed, apart from moving from the last state of a given level to the first state of the next one, which is performed automatically once the last state in a level is reached.

The point in the escape rooms with this layout is to move on by taking a series of tests, composed of a number of questions out of a question bank related to a given didactic unit. Those tests may include multiple choice questions, filling in the blank or calculated questions with random variables. The number of questions per test may depend on the top marks expected and the score assigned to each item. For instance, a 10-question test where each one is worth 1 point will get a top mark of 10.

Regarding its implementation, a particular class session can be devoted to carrying out the escape room, where students may get together in groups, or otherwise, they may go on their own, according to the teacher's decision. The goal in any case is to take a series of tests as fast as possible in order to advance in the scenario proposed, as the time to get it all done is constrained. This way, students have to apply all the knowledge they acquired from the didactic units, either during the class sessions or studying by themselves. Hence, this approach for the escape room could be seen as a bunch of exams to be carried out either in groups or on an individual basis.

Sticking to the Spanish grading system (Polytechnic University of Valencia – Office of International Exchange Programmes, 2020), the scores range from 0 to 10, where 5 is the passing grade. In this scoring framework, the initial state in each level p_j is state $s_{j,0}$ and it contains six more states, ranging from $s_{j,1}$ all the way to $s_{j,6}$. Besides, upon reaching this last state within the level p_j, namely $s_{j,6}$, that level gets cleared, and in turn, the next level is started out, expressed as a state $s_{j+1,0}$, which belongs to the level p_{j+1}. Furthermore, clearing the last level implies the end of the game for the group doing so in the case of a team competition, or otherwise, for the student doing so in the case of an individual competition.

As stated above, the procedure of the escape room requires all participants, either groups or individual students, to take a series of 10 randomly chosen question quizzes out of the chapter question bank. In order to minimize the risk of question repetition among participants, each question bank should be large enough, whilst it should also include a number of questions with random variables. When a participant completes a given test, if the outcome is lower than 5, then the participant must repeat the test over again so as to try and get better marks. Otherwise, if the outcome is equal or higher than 5, then the participant can hop forward to further states according to Table 1, where just the integer part of a possible score is shown. Hence, the higher the score in a test, the further the movement is made in the escape room.

However, a restriction on the number of hops away must be imposed, as its maximum number depends on how far the completion of a level is. This way, if a participant is located in the initial state of a level, then three different situations may arise after taking the first quiz of that level. To begin with, if the maximum score is obtained, then 6 hops away are achieved. Hence, the participant gets to the last state of that level, thus it is automatically promoted to the initial state of the next level. Otherwise, if a mark ranging from 5 to 9 is obtained, then a number of hops is achieved according to Table 1. Furthermore, if a mark lower than 5 is obtained, no moves are gained, as that mark is lower than the passing grade.

On the other hand, if a participant is not located in the initial state of a level, the same approach applies, even though the last state of that level is closer. As a consequence, the last state can be achieved with a lower score than the top grade, according to how far the participant is from it. In this sense, it must be considered that the excess of marks when getting to the final

Table 1. Hops away regarding the test score (just its integer part)

Score	0–4	5	6	7	8	9	10
Hops away	0	1	2	3	4	5	6

Source(s): Authors' own work

step of a level gets discarded. Otherwise, a participant remains at the same level as long as it does not reach its last state.

Once a test is done and the movement associated with the score achieved has been accounted for, a participant can begin a new test as soon as possible in order to advance through the different states $s_{j,i}$ and levels p_j. This way, the faster a test is completed, the quicker a new test can be started, thus the participant can move more quickly throughout the escape room (Veldkamp *et al.*, 2020). In any case, the goal is to clear all levels as soon as possible, and the participant doing so will be the winner of the escape room (Alabdulaziz, 2023).

Regarding the final marks assigned to each participant in the escape room, it is done according to the time elapsed to complete it. For instance, top marks may be assigned to all participants who complete the escape room within the 60% of the session duration, whilst passing marks may be associated with the participants who clear it by the end of the session, which stands for the 100% of the session duration. This way, different passing grades may be linked to completing the escape room within 70%, 80%, 90% or 95% of the session duration, whereas not finishing the escape room may result in failing marks.

Therefore, referring to the Spanish grading system, a score of 10 would be assigned to the participants finishing the escape room before the 60% of the session duration. Then, a score of 9 would be granted to fulfill the escape room before the 70% of the session duration, a grade of 8 before the 80%, a grade of 7 before the 90%, a grade of 6 before the 95%, and eventually a grade of 5 before the 100%. Otherwise, grades under the cutoff mark would be assigned according to the number of levels completed when the session is done.

The description proposed could easily be extrapolated to any other grading system by first changing the top mark and the passing mark, and in turn, adapting the number of states within each level to the grades obtained. For this purpose, a mathematical formalization is introduced to provide a more abstract framework that can subsequently be particularized to fit another grading system. This point is going to be developed in the next subsection.

2.2 Formal representation

The schematic diagram proposed above for the Spanish grading system has 7 states within each level, namely the initial state 0 and also states from 1 to 6. The reason for the latter is that the top mark is 10, the passing mark is 5, and the step between two consecutive marks is taken as 1. This way, the set of passing marks considered therein is isomorphic to the segment of natural numbers from 5 up to 10, as depicted in Table 1. Hence, the number of states in a level may be viewed as the difference between the top mark and the passing mark, divided by the step taken. However, an extra unit must be added to that difference in order to count the lower value of the range. Besides, another extra unit is needed to account for the initial state of a level. Consequently, the number of states in a level proposed in the previous subsection was $(10 - 5) / 1 + 2 = 7$.

This scheme can be easily generalized so as to adapt it to any grading system by just modifying the top grade, the passing grade and the step taken. This type of modification would result in keeping the evenly distributed distances between states because the step is constant, whereas adopting a variable step could result in unevenly distributed distances. Therefore, this scheme may be tailored in order to adapt it to different types of grading systems. On the other hand, the number of steps must always be a natural number, the result of the aforementioned division should be truncated in order to achieve so.

Sticking to the case of a constant step, a formal representation for a generic grading system is presented in Figure 1, where α is set to be the minimum grade to pass an exam, whilst β is assigned to the maximum available grade, and δ is associated with the step. Taking this into account, σ indicates the number of states within a level, where the value of the step taken between any pair of neighboring states is constant, thus resulting in (1). Obviously, if the step considered is $\delta = 1$, it effectively simplifies the calculation of σ, such that $\sigma = \beta - \alpha + 2$, which is the case for the number of states presented in the previous subsection.

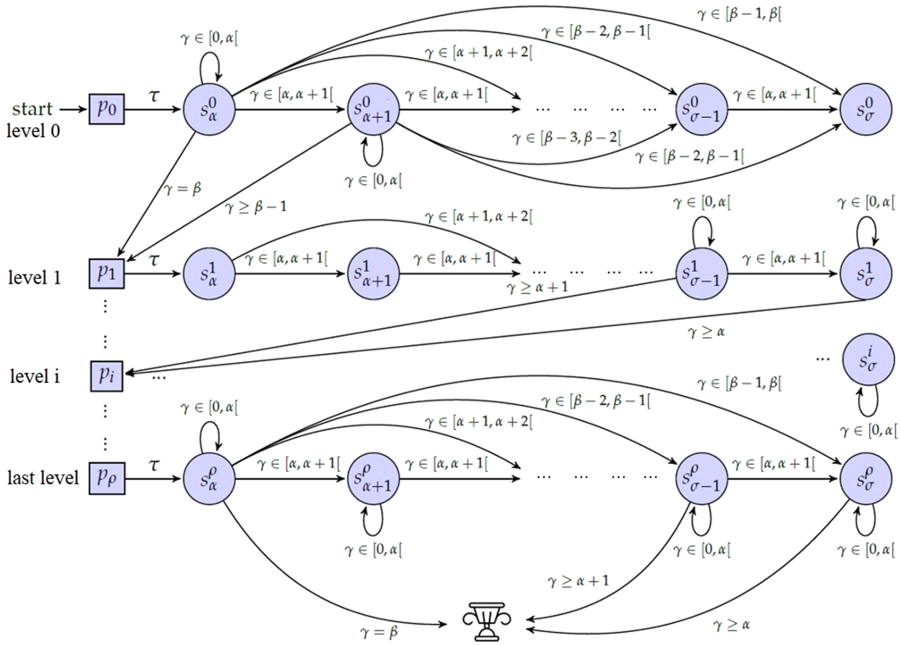

Figure 1. Flow chart for the behavior of an escape room proposed with a generic grading system. Source: Authors' own work

$$\sigma = \left\lfloor \frac{\beta - \alpha}{\delta} \right\rfloor + 2 \tag{1}$$

A similar expression could be worked out in order to calculate the number of hops away due to a given score, given by n. This way, such an expression would map the score obtained in a particular test to the number of hops away achieved due to that score, thus mirroring the outcome of Table 1. In order to do so, the top grade in the former expression should be substituted with the actual grade of a particular test, namely γ, as in expression (2). However, it does not apply if γ does not reach the passing grade.

$$n = \left\lfloor \frac{\gamma - \alpha}{\delta} \right\rfloor + 2 \tag{2}$$

Nonetheless, in order to account for a null movement if the passing grade is not reached, then a piecewise function can be defined to state so. This is shown in expression (3).

$$n = \begin{cases} \left\lfloor \frac{\gamma - \alpha}{\delta} \right\rfloor + 2 & if\ \gamma \geq \alpha \\ 0 & if\ \gamma < \alpha \end{cases} \tag{3}$$

Regarding Figure 1, each state bears a superindex, whose role is to carry the level it belongs to, as well as a subindex, whose role is to convey its state identifier within that level. Furthermore, $\rho + 1$ stands for the number of levels involved in the escape room, those ranging from 0 all the way to ρ. Moreover, γ stands for the grade obtained in a given test, which is represented as a continuous variable, thus assigning a range of qualifications to the movement towards a certain state. Nonetheless, this could also be applied to discrete variables, such as the natural numbers, where a single qualification or a range of them could be assigned to a certain movement.

Furthermore, it is to be said that not all arrows are shown in the picture, as each state actually has arrows pointing out to the rest of its neighbors ahead within its own level. Hence, just some of the arrows have been displayed for clarity purposes. In addition to it, they also have another arrow pointing out to the initial state of the following level. Therefore, for clarification purposes, only a handful of arrows have been depicted therein. Also, loopback arrows are located in each state in order to visualize the effect of failing tests, which results in remaining in the same position, although only some of them have been shown.

Additionally, each level has an extra initial point, which is exhibited as a square identified with p_i, where i is the level identifier. The aim of those points is to concentrate all arrows coming out of each state from the previous level, which accounts for clearing the previous level and getting into the next one. The role of such points is to act as a hub, as they all have an arrow pointing out to the first state of their own level. This arrow is considered a silent transition τ, meaning that a participant reaching any initial spot within a level is automatically redirected to the first state of such a level. Besides, all states in the last level have arrows pointing out to the winning cup, which stands for the end of the escape room (Sipone *et al.*, 2024).

In summary, the proposed formalization provides three key features: portability across grading schemes, implementability and reproducibility. With regard to portability, the formal model described above has been described for a generic grading system. However, it can easily be adapted to any specific grading system by substituting the generic variables defined in the model with the actual variables used in that particular grading system, regardless of whether numerical scores or letter grades are employed.

With respect to implementability, two perspectives can be distinguished depending on whether the grading scheme uses numerical scores or letter grades. The former case is simpler, as the specific values of the top score and the passing grade are used, whereas the step value can be defined according to the desired number of states per level. Furthermore, unequal values can be assigned to different forward hops, which also provides a useful solution when the value obtained for σ is not a natural number. Likewise, the latter case requires assigning a correspondence between letter grades and forward hops, along with establishing a step value to obtain the desired number of states per level. Additionally, in both cases, a tailor-made passing grade can be defined in order to raise or lower the threshold required to progress within the escape room, thus allowing adaptation to different circumstances.

Regarding reproducibility, the formal model relies on basic arithmetic operations, ensuring that identical analyses performed on the same scores yield the same outcomes. This way, verifiability can be ensured, transparency is achieved, and confidence in the integrity of the method is strengthened. Furthermore, the formal model can be considered as both sound, given the structural correctness of its deductive system, and complete, in that all true statements expressible within the system are provable.

2.3 Use case of a multistage linear path design

This multistage linear path model has been used to set up an escape room as an evaluation tool in a course on the basics of networking. We established four levels, where each of them was dedicated to a different layer within the Open Systems Interconnection (OSI) model. It should be noted that the OSI model is a conceptual framework composed of seven layers, designed to distribute the roles of the different protocols involved in network communication. From a practical perspective, the lower four layers and the top layer are the most important. Hence, the multistage linear-path design allows each branch of the escape room to be dedicated to a specific layer, such that all questions included in a test within a given branch are devoted exclusively to that layer. Accordingly, a separate question pool of 200 questions was created for each of the four lower layers, with Level 1 corresponding to Layer 1, and so forth.

As stated above, progression in the escape room depends on the score achieved in a series of ten-question tests, whose questions are randomly selected at runtime from the pool associated with the branch in which a given player is located. In this way, students face a series of tests related to each of the four lower layers in an ordered manner. Moreover, since ten questions are drawn from a pool of 200 questions, the probability of encountering the same question in two different tests is 5%, which can be considered a low-probability event. The probability is computed assuming uniform random selection of questions with replacement across attempts, which can be seen as sufficient to justify the low likelihood of question repetition in the proposed design.

As shown in Table 1, each test score is associated with a specific movement within a given branch of the escape room, with the objective of reaching the final state of that branch as quickly as possible and progressing to the next branch. Accordingly, each team must repeatedly take ten-question tests corresponding to the current branch, all of which focus on a particular layer of the OSI model; therefore, higher scores allow the branch to be cleared more rapidly. Furthermore, the fact that the escape room is played on a team basis encourages collaborative problem-solving among team members. The escape room is ultimately completed once all branches have been cleared sequentially, with the winning team being the one that finishes in the shortest time.

The use case scenario can be regarded as a proof of concept demonstrating the practical feasibility of implementing the proposed multistage linear-path escape room design. In this sense, it can be adapted to any type of course or to a specific part of a course. The time required to complete the activity depends on the difficulty of the proposed questions and the number of levels selected. For instance, the reported use case was based on a STEM course comprising four levels and was implemented within a 90-min session, during which most teams managed to complete the escape room. Alternatively, a three-level deployment could be implemented within a 60-min session, even though teams might experience greater difficulty in completing the activity due to tighter time constraints.

Regarding the resources required, a question pool is needed for each level, with at least 100 questions per pool to keep the probability of encountering the same question in two different tests at 10%. Increasing the pool size to 200 questions reduces this probability to 5%, thereby making question repetition across tests a low-probability event. With respect to teacher training, no specific training is required, as instructors act primarily as facilitators of the escape room while students assume the active role. Moreover, with regard to long-term pedagogical impact, this type of activity supports continuous improvement, deeper student understanding, stronger links between theory and practice, professional development and an action-oriented approach (Rogti, 2021).

Focusing on the nature of the questions within each question bank, we attempted to design all questions with a similar level of difficulty, so that the overall difficulty remained reasonably equivalent across the various trials. Each question bank included a mix of MCMA questions (Multiple Choice, Multiple Answers), MCSA questions (Multiple Choice, Single Answer), fill-in-the-blank questions and calculated questions. Furthermore, some MCMA and MCSA questions were designed to mirror the fill-in-the-blank and calculated questions, helping to ensure a comparable level of difficulty across the question sets.

2.4 Methodological limitations

The use case described above has been implemented for the first time and, as such, entails some methodological limitations. In this context, the pilot course in which this escape room was implemented for assessment purposes included 60 students, all of whom participated. Hence, the first limitation concerns the absence of a simultaneous control group, which limits the ability to conduct direct comparisons of the observed outcomes. To address this issue, the cohort could have been evenly split into a test group and a control group, with the former receiving the new assessment scheme and the latter following a traditional assessment scheme, thereby serving as a baseline. However, we considered it more important to assess all students in the same way in order to avoid potential unfairness. Nonetheless, we considered using a similar course delivered in the previous year, with a comparable number of students, as a control group, even though this approach may introduce time-related bias.

Another limitation concerns the lack of internal validation of the escape room, which can be examined from both a structural and a content-related perspective. With respect to the structure, formal model-checking techniques could be applied to the formal model in order to assess its liveness and safety properties. Safety ensures that undesirable states are never reached, whereas liveness guarantees that desirable states are eventually attained. In the proposed escape room, progression is strictly forward-oriented; hence, contestants will eventually obtain the required scores to complete the activity, ensuring that they cannot remain indefinitely stuck. With regard to the content of the four question pools created, the appropriateness of the questions included in each pool was validated by all instructors teaching the course.

On the other hand, another limitation concerns the small sample size, as well as the lack of randomization. In this sense, the escape room was implemented in a single pilot course, and all enrolled students were included. Consequently, the sample size could not be larger, and no random selection was possible. Another limitation concerns the potential influence of instructor or contextual effects on students' learning outcomes or classroom experience. Both effects can be considered negligible, as all students in the course delivered in the current year (the test course) and in the previous year (the control course) were taught by the same instructors. Moreover, both cohorts consisted of second-year students and were therefore already adapted to the campus dynamics.

Finally, we acknowledge that the unit of analysis adopted in the inferential statistics may be considered somewhat controversial, as team-based outcomes were obtained in the current course (experimental group), whereas student-level outcomes were collected in the previous course (control group). To address this issue, the inferential results are presented using a dual approach. On the one hand, the team scores obtained in the current course were assigned to each corresponding team member in order to conduct the analysis at the student level, treating observations as independent. On the other hand, the distribution of student scores from the previous year was grouped into clusters of four students, allowing the inferential analysis to be conducted at the team level using a clustered modelling approach.

3. Results

This type of multistage escape room was used as an evaluation tool in a course devoted to the basics of networking in a university degree. That course had 60 students registered in the present academic year, where all of them were aged from 18 to 24 years old. On the other hand, the same course was run in the last academic year, with the same number of students registered and the same range of ages.

In the current year, an evaluation was undertaken by means of a multistage linear escape room, where students were randomly organized in groups of four students. Regarding the group formation, it is to be taken into account that some of the students in the current year are those who did not manage to pass the course in the last academic year. However, this fact did not imply anything about the chances of a given student performing better or worse in an exam, as all of them had access to the same materials, no matter if they were first-time learners in the course or otherwise. Therefore, the potential impact on results related to group formation was irrelevant.

However, a traditional evaluation was carried out last year by means of a written exam. The outcome attained with those written exams in the prior academic year is going to be compared with the results obtained with the evaluation through the escape room in the current academic year. In this sense, the average marks obtained in the last year were 6.61 out of 10, where the latter is the top mark in the Spanish grading system, whereas the success rate in the last year was 41 out of 60 students.

As the escape room was held as a team effort, where the 60 students registered in the course were randomly distributed in groups of 4, then there were $60/4 = 15$ teams taking part in the escape room. The distribution of the scores obtained by each team according to the time they took to clear the escape room is given in Table 2.

On the other hand, the distribution of scores per student is displayed in Table 3, considering that each team is composed of 4 members.

From those results, it is possible to apply descriptive statistics in order to find out the main centralization and dispersion measurements, as shown in Table 4. It should be noted that the reported measures of central tendency and dispersion yield the same values when computed at the team level and at the student level, as sample size does not affect any of the calculations involved.

Table 5 summarizes the relevant descriptive statistics extracted from both academic years, where Sample A corresponds to the previous year (the control group), and Sample B corresponds to the current year (the test group).

In addition, confidence intervals and error measures for the means can be derived from the descriptive statistics presented on a team basis, as displayed in Table 6.

Alternatively, confidence intervals and error measures for the means can be derived from the descriptive statistics presented on an individual basis, as displayed in Table 7.

Furthermore, an inferential statistical study has been undertaken so as to establish if the improvements achieved are statistically significant. Using the relevant data obtained from the

Table 2. Distribution of the scores obtained by each team

Score	0	1	2	3	4	5	6	7	8	9	10
Number of teams	0	0	0	1	2	1	0	1	4	5	1

Source(s): Authors' own work

Table 3. Distribution of the scores obtained by each student

Score	0	1	2	3	4	5	6	7	8	9	10
Number of students	0	0	0	4	8	4	0	4	16	20	4

Source(s): Authors' own work

Table 4. Descriptive statistics of the scores obtained per team and per student

Type	Statistic	Value
Centralization	Average	7.33
	Mode	9
	25th percentile	5
	50th percentile	8
	75th percentile	9
Dispersion	Variance	4.62
	Standard deviation	2.15
	Coefficient of variation	0.29

Source(s): Authors' own work

Table 5. Most relevant descriptive statistics of both samples per team and per student

Sample identifier	Sample size per team (N_T)	Sample size per student (N_S)	Average (M)	Standard deviation (SD)
A	$N_1 = 15$	$N_1 = 60$	$M_1 = 6.61$	$SD_1 = 1.65$
B	$N_2 = 15$	$N_2 = 60$	$M_2 = 7.33$	$SD_2 = 2.15$

Source(s): Authors' own work

Table 6. Confidence intervals and error measures for the means on a team basis

	Error measure	68% confidence interval	90% confidence interval	95% confidence interval	99% confidence interval
Sample A	0.42	[6.19–7.03]	[5.91–7.31]	[5.78–7.44]	[5.52–7.70]
Sample B	0.56	[6.78–7.89]	[6.42–8.25]	[6.25–8.42]	[5.90–8.77]

Source(s): Authors' own work

Table 7. Confidence intervals and error measures for the means on an individual basis

	Error measure	68% confidence interval	90% confidence interval	95% confidence interval	99% confidence interval
Sample A	0.21	[6.40–6.82]	[6.26–6.96]	[6.19–7.03]	[6.06–7.16]
Sample B	0.28	[7.06–7.61]	[6.88–7.79]	[6.79–7.88]	[6.62–8.05]

Source(s): Authors' own work

descriptive statistics, a *t*-test has been carried out in order to check out if there is a statistically significant difference between both samples on a team basis, as shown in Table 8.

Alternatively, the calculation of a *t*-test and the *p*-value has also been undertaken on an individual basis, as displayed in Table 9.

At this point, the effect size achieved has been found, as shown in Table 10. It should be noted that the means and standard deviations are identical when computed at the team level and at the student level for each sample. Moreover, since the sample sizes are the same at the team level and the same applies at the student level, the effect size remains unchanged in both cases.

Table 8. Inferential statistics for the distribution of scores in each year on a team basis

Variable	Value
Value given by the *t*-student distribution	1.04
Degrees of freedom (df)	28
p-value	0.31
s-value (α value)	0.05
Interpretation	*p*-value > *s*-value: not statistically significant

Source(s): Authors' own work

Table 9. Inferential statistics for the distribution of scores in each year on an individual basis

Variable	Value
Value given by the *t*-student distribution	2.06
Degrees of freedom (df)	118
p-value	0.04
s-value (α value)	0.05
Interpretation	*p*-value < *s*-value: statistically significant

Source(s): Authors' own work

Table 10. Effect size achieved at both the team level and the student level

Variable	Value
Difference of means: $\lvert M_2 - M_1 \rvert$	0.72
Pooled standard deviation (Pooled SD)	1.92
Cohen's d (effect size)	0.38
Interpretation	0.20 < 0.38 < 0.50: small to medium effect

Source(s): Authors' own work

Besides, the sample size needed to achieve the effect size has been calculated, as seen in Table 11. As a side note, since the effect size is the same at both the team level and the individual level, the required sample size is also the same in both cases.

After finishing the escape room, the level of engagement was measured by means of the ISA engagement scale (Soane *et al.*, 2012). This is a specific construct composed of 9 standard questions, which are grouped in 3 different dimensions: intellectual, social and affective. Actually, the word ISA comes from the acronym of those 3 dimensions involved. Table 12 exhibits the three questions associated with each dimension.

Additionally, each of the 9 standard questions overall are to be rated in a 7-point Likert-type scale, where a rate of 1 stands for "fully disagree," a rate of 2 stands for "disagree," a rate of 3 stands for "partially agree," a rate of 4 stands for "neither agree nor disagree," a rate of 5 stands for "partially disagree," a rate of 6 stands for "agree," and finally, a rate of 7 stands for "fully agree".

This way, the level of engagement for each dimension corresponds to the average value of all ratings for the 3 questions related to that dimension. This way, 3 levels of engagement are

Table 11. Sample size needed to obtain the effect size achieved at both the team level and the student level

Variable	Value
Cohen's d	0.38
Alpha value	0.05
Beta value	0.20
Sample size needed	112
Actual sample size (at team level)	15
Actual sample size (at student level)	60
Interpretation	Actual sample size (either at team level or at student level) < needed sample size: not sufficient to detect the effect size achieved

Source(s): Authors' own work

Table 12. Questions within the ISA engagement scale

Dimensions	Items
Intellectual	Q1: I focus hard on my work
	Q2: I concentrate on my work
	Q3: I pay a lot of attention to my work
Social	Q4: I share the same work values as my colleagues
	Q5: I share the same work goals as my colleagues
	Q6: I share the same work attitudes as my colleagues
Affective	Q7: I feel positive about my work
	Q8: I feel energetic in my work
	Q9: I am enthusiastic in my work

Source(s): Soane *et al.* (2012).

obtained, referred to as 3 different points of view, namely, intellectual, social and affective. Eventually, the overall level of engagement is also obtained by calculating the average of all three dimensions, which accounts for the overall average of all ratings obtained in the construct. Therefore, as the rate of 6 stands for "agree," the minimum measurement expected in each dimension and overall is 6 (Sidharta, 2019). Table 13 displays the average scores obtained in the ratings of the questions related to each dimension, along with the overall average scores.

It should be noted that the ISA engagement scale is a standardized instrument for assessing engagement levels and has been widely used across different contexts, including workplace and educational settings. This instrument was validated at the time of its development; therefore, there was no need to conduct an additional content-validation procedure such as Aiken's V test.

The construct was administered at the end of the class session in which the escape room was implemented, with the aim of assessing students' engagement during the activity. Specifically, it was delivered as a computer-based questionnaire, in which students were asked to provide ratings for each of the nine items included in the ISA engagement scale.

Moreover, the reliability of the engagement instrument was assessed using Cronbach's alpha, which was computed for the overall scale as well as for each specific dimension. Table 14 presents the results obtained in all cases.

Regarding the academic performance and success rate achieved in both academic years, Table 15 compares those variables and displays the percentage of increase obtained in the present course (test group) with respect to the previous course (control group).

Additionally, confidence intervals and error measures for the pass rates can be derived from the descriptive statistics presented on a team basis, as displayed in Table 16.

Alternatively, confidence intervals and error measures for the pass rates can be derived from the descriptive statistics presented on an individual basis, as displayed in Table 17.

Table 13. Level of engagement obtained per dimension and overall

Type of engagement	Score in the ISA construct
Intellectual level of engagement	6.51
Social level of engagement	6.67
Affective level of engagement	6.72
Overall level of engagement	6.63

Source(s): Authors' own work

Table 14. Reliability of the engagement instrument assessed through Cronbach's alpha

	Intellectual dimension	Social dimension	Affective dimension	Overall instrument
Cronbach's Alpha	0.90	0.84	0.70	0.71

Source(s): Authors' own work

Table 15. Comparing the academic performance and success rate in both years

Metric	Prior year	Present year	Ratio	% increase
Academic performance	6.61 out of 10	7.33 out of 10	$\frac{7.33}{6.61} = 1.11$	11%
Success rate	41 out of 60	48 out of 60	$\frac{48}{41} = 1.17$	17%

Source(s): Authors' own work

Table 16. Confidence intervals and error measures for the pass rates on a team basis

	Error measure	68% confidence interval	90% confidence interval	95% confidence interval	99% confidence interval
Sample A	0.12	[0.56–0.80]	[0.49–0.88]	[0.45–0.92]	[0.37–0.99]
Sample B	0.10	[0.70–0.90]	[0.63–0.88]	[0.60–1.00]	[0.53–1.07]

Source(s): Authors' own work

Table 17. Confidence intervals and error measures for the pass rates on an individual basis

	Error measure	68% confidence interval	90% confidence interval	95% confidence interval	99% confidence interval
Sample A	0.06	[0.62–0.74]	[0.58–0.78]	[0.57–0.80]	[0.53–0.84]
Sample B	0.05	[0.75–0.85]	[0.72–0.88]	[0.70–0.90]	[0.67–0.93]

Source(s): Authors' own work

4. Discussion

Regarding the results shown in Table 2, 12 out of 15 teams were able to finish the escape room in time. Those are the teams getting a score equal to or higher than 5, which is the passing grade in the Spanish grading system. Hence, considering that each team was formed by 4 team members, it accounted for 48 out of 60 students who got to finish the escape room in time, as depicted in Table 3.

Since grades in the current course were obtained on a team basis, whereas grades in the previous course were collected at the individual level, a dual approach was adopted in the statistical analysis. Under the team-level approach, the distribution of individual scores from the previous year was grouped into clusters matching the team size, enabling comparison through a clustered modeling framework. Under the student-level approach, each team's score was assigned to all corresponding team members, allowing the analysis to be conducted at the individual level.

With respect to the results exhibited in Table 4, a descriptive statistical analysis was applied to the outcome by calculating some key figures in descriptive statistics. It should be remembered that the descriptive statistics obtained do not depend on whether the calculations are performed at the team level or at the student level, since all teams consisted of the same number of students. Consequently, the only difference between the two approaches is the sample size, which does not enter into any of the descriptive measures reported.

Focusing on the centralization measurements, the average value obtained was 7.33 out of 10, which is considered a remarkable grade in the Spanish grading system. Besides, the mode was 9 out of 10, which is considered an outstanding grade in that system. Moreover, the values of the quartiles are 5 for the first one, 8 for the second one, also known as the median, and 9 for the third one, which are quite high marks.

Centering on the dispersion measurements, the variance of the results obtained was 4.62, whereas the standard deviation was 2.15. This led to a coefficient of variation of 0.29, as a result of dividing the average by the standard deviation. It is to be noted that values for the coefficient of variation lower than 0.30 suggest moderate variability, and this is the case for the results achieved.

Subsequently, Table 5 summarizes the most relevant descriptive statistics, namely the mean and standard deviation, for both courses, along with the sample sizes at the team and student levels. Table 6 then reports the error measures and confidence intervals for the means computed at the team level, whereas Table 7 presents the corresponding results at the individual level. With respect to the error measures, the values obtained at the individual level are half those computed at the team level. This is because the sample size in the former is four times larger, and the error measure depends inversely on the square root of the sample size. With regard to the selected confidence intervals, they approximately correspond to ±1 standard deviation (68%), ±1.66 standard deviations (90%), ±2 standard deviations (95%) and ±2.5 standard deviations (99%), under the assumption that the data obtained follow a normal distribution.

Sticking to inferential statistics, Table 8 displays the results of the *t*-test and the *p*-value on a team basis. The *t*-value obtained is 1.04, whereas the corresponding *p*-value is 0.31. Consequently, the team-level results may not be considered statistically significant, since the *p*-value exceeds the selected significance level ($\alpha = 0.05$).

On the other hand, Table 9 exhibits the outcome of the *t*-test and the *p*-value on an individual basis. The *t*-value obtained is 2.06, whereas the corresponding *p*-value is 0.04. Consequently, the team-level results may be considered statistically significant, since the *p*-value is lower than the selected significance level ($\alpha = 0.05$). This might imply that the differences found in both samples are unlikely to have happened due to chance alone, so it allows to suggest the statistical significance of the reported improvements on a student basis.

Additionally, the effect size was calculated in Table 10 with the means and standard deviations of the distribution of marks collected in both courses, thus leading to a value of 0.38. This value is located around the middle point of the interval between 0.2 and 0.5, where the former accounts for a small effect and the latter represents a medium effect. Actually, a medium effect of 0.5 stands for an effect which is visible to the naked eye, considering a careful observer, whereas a small effect of 0.2 stands for an effect which is noticeably smaller than the medium one. It should be noted that the effect size is identical at both the team and individual levels, since the variables used in both calculations take the same values.

Once the effect size achieved was found out, the sample size needed to achieve that effect size was calculated in Table 11 by considering an alpha value of 0.05 and a statistical power of 0.80, which accounts for a beta value of 0.20. The outcome of the sample size needed with those inputs was 112, which is nearly twice the number of participants in the courses and nearly eight times the number of teams involved. Therefore, it suggests that the actual number of students in the courses was not sufficient to detect the effect size achieved on an individual basis, whereas the number of teams involved was not sufficient either. Hence, further research should be carried out with a larger number of students and teams.

Therefore, analyzing student-level observations as independent despite the team-based grading used in the escape-room may affect the *t*-test outcome and the associated *p*-value due to the potential violation of the independence assumption. However, the estimated effect size and the corresponding required sample size remain unchanged under both assumptions. Likewise, the computation of common descriptive statistics does not differ between the two approaches. In other words, the only statistical measures affected by adopting a team-level versus an individual-level approach are those that depend on sample size.

It should be noted that, although the sample size represents a limitation, the proposed model could be adapted to larger classes by organizing students into multiple parallel teams that follow the same activity structure, allowing the methodology to scale without substantially modifying the design. In such cases, digital platforms or learning management systems could be used to distribute materials, manage progress and coordinate interactions between teams. Furthermore, this approach could be extended to online learning environments by implementing the activities through collaborative tools and virtual breakout rooms, thus enabling students to engage in the same sequence of tasks remotely while preserving the collaborative and problem-solving elements of the methodology.

On the other hand, with regard to the outcome of the ISA engagement scale, it is to be noted that it was originally created for working environments, even though it has been exported to other fields, such as the education area (Nwachukwu and Osa-Izeko, 2022). In this case, the outcome displayed in Table 13 reveals values of engagement level higher than 6, which accounts for ‘agree’, in all three dimensions considered and overall. It should be noted that a value of 6 is the minimum expected score to account for a high level of engagement, according to the mapping between rates and meanings for the ISA construct established above (Mañas-Rodríguez *et al.*, 2016). Actually, the results obtained were all above 6.5 in all cases, such that they were even closer to 7 than 6, where the value of 7 stands for “fully agree.” Hence, it seems that the engagement level was really high according to those figures.

Furthermore, looking deeper into those results, it appears that the intellectual engagement was a bit lower than the social and affective ones. This could be explained by the fact that it seems that students’ engagement was more strongly driven by getting connected to their peers and how they feel about it than by the knowledge they acquire. Analogously, this point could also be applied to explain that the social engagement was slightly lower than the affective one, as it seems that the connectivity with others was a bit less important than the feelings about it.

Focusing on the factors driving the engagement in the three dimensions during the escape room, the intellectual engagement was mostly driven by the will to get a good performance in the escape room due to the competitive environment among the teams in order to beat the rivals. Moreover, the social engagement was mostly driven by the desire to be not only part of a group but also fully integrated with the rest of the components of the group, which was part of the team spirit brought by the competition. Besides, the affective engagement was mostly driven by the feeling involving the gaming environment of the competition, which allowed the creation of bonds among the teammates.

In addition, the reliability of the ISA engagement scale was assessed using Cronbach’s alpha, which was computed for the overall instrument as well as for each specific dimension. According to Table 14, the overall Cronbach’s alpha was 0.71, which is generally regarded as acceptable (0.70–0.80). Furthermore, the intellectual dimension yielded an alpha of 0.90, suggesting excellent reliability (0.90–1.00), whereas the social dimension reached 0.84, suggesting good reliability (0.80–0.90). Finally, the affective dimension obtained a value of 0.70, which suggests acceptable reliability.

Regarding academic performance and success rate, Table 15 compares the outcomes achieved in both academic years. Those results display an increase in academic performance of 11% between the course run in the last academic year (control group) and the course run in the present academic year (test group). Likewise, the results exhibit an increase in success rate of 17% between both courses.

Additionally, Table 16 then reports the error measures and confidence intervals for the pass rates computed at the team level, whereas Table 17 presents the corresponding results at the individual level. With respect to the error measures, the values obtained at the individual level are half those computed at the team level because the individual-level sample size is four times greater, and the standard error depends inversely on the square root of the sample size.

Eventually, it has been widely reported in the literature that the use of active learning approaches in STEM education leads to an increase in students' performance (Freeman *et al.*, 2014). Specifically, an increase of around 15% has been reported for academic performance (Hacisalihoglu *et al.*, 2018), whereas an increase of around 20% has been reported for success rate (Nurbavliyev *et al.*, 2022). Therefore, a comparison has been made between the results obtained in the last academic year and the results obtained in the present academic year at both the team level and the individual level.

Therefore, the 11% increase in academic performance observed in the current course relative to the previous one is broadly consistent with figures reported in the literature, which are around 15%. Likewise, the 17% rise in success rate compared to the previous course is close to the approximately 20% improvement documented in prior studies.

5. Conclusions

In this paper, a multistage linear path structure is proposed for educational escape rooms, where each stage, also referred to as a level, needs to be cleared before moving to the following one. This way, each level is considered as an independent entity, thus allowing to dedicate each one to the knowledge, skills and abilities learned in a particular didactic unit within the curriculum of a given course.

The structure presented is first described as an informal representation with the specific characteristics of the Spanish grading system, which in turn gets generalized as a formal representation in order to extend it to different grading systems. In this sense, a flow chart of the behavior of the escape room is provided, as well as a couple of algebraic expressions about it.

Afterwards, the results obtained in the present academic year with the escape room as an active evaluation tool are presented, which in turn are compared with the results achieved in the last academic year with a traditional written exam. In summary, results showed an increase in academic performance of 11%, as well as an increase in success rate of 17%, which are in line with those quoted in the literature.

However, it should be noted that outcomes in the current academic year (the test group) were obtained on a team basis, whereas outcomes in the previous academic year (the control group) were collected at the individual level. Therefore, the statistical analysis was conducted using two units of analysis: the team level and the student level. Under the student-level approach, each team's score from the current course was assigned to the corresponding team members, whereas the distribution of individual scores from the previous course was grouped into clusters matching the team size. As a result, the number of teams and students considered in both cohorts was equivalent, which facilitates comparability in the statistical analysis.

Furthermore, an inferential study was carried out on the distribution of scores obtained in both courses at both the team and individual levels. At the team level, the resulting t-value was 1.04, and the corresponding p-value was 0.31. Since the p-value is greater than the selected significance level ($\alpha = 0.05$), it appears that the improvement cannot be considered statistically significant. On the other hand, at the student level, the resulting t-value was 2.06, and the corresponding p-value was 0.04. Since the p-value is lower than the selected significance level ($\alpha = 0.05$), it suggests that the improvement may be statistically significant.

Besides, the effect size achieved was also calculated, leading to a value of 0.38 at both the team level and the student level. This value is located around the middle of the range between 0.2 and 0.5, where the former represents a small effect size and the latter stands for a medium effect size. The latter accounts for an effect which is visible to the naked eye, considering a careful observer, whereas the former stands for a noticeably smaller effect than the medium one.

Moreover, the sample size needed to achieve this effect size was also calculated. In order to do so, an alpha value of 0.05 was established, as well as a statistical power of 0.80, which accounts for a beta value of 0.20. The result obtained for the sample size needed was 112 in both the team level and the student level, which is nearly eight times the number of teams involved, namely 15, and nearly twice the learners attending the courses, namely 60. Hence, it appears that the number of students was not sufficient to detect the observed effect size, and further research should be conducted with a larger group of learners. The same limitation applies to the number of teams considered in the analysis.

Additionally, the level of engagement when taking the escape room was measured by means of the ISA engagement scale. This measurement takes into account three dimensions, such as intellectual, social and affective, where all those dimensions involved achieved an outcome greater than 6. Such results led to an overall amount also greater than 6, which suggests a high level of engagement when carrying out the escape room. Therefore, it seems that students experienced quite a high engagement level during the escape room.

Finally, based on the results obtained, the research question posed in the introduction can be answered positively, as it was possible to implement an escape-room-based assessment to evaluate a course while observing improvements in academic performance and increased learner engagement. However, it should be acknowledged that the observed trends cannot be taken as definitive evidence of causal impact resulting from the proposed model's implementation, but rather as indicative of a plausible mechanism. Nonetheless, further research should be conducted with a larger group of students in order to more thoroughly evaluate the proposed model.

Author contributions
All authors contributed equally in this manuscript. All authors have read and agreed to the published version of the manuscript.

Institutional review board statement

This study was conducted in accordance with the Declaration of Helsinki. No approval by the Institutional Ethics Committee was necessary, as all data were collected anonymously from capable, consenting adults. The data are not traceable to participating individuals. The procedure complies with the General Data Protection Regulation (GDPR), currently in force in the European Union.

Informed consent statement

Informed consent was obtained from all subjects involved in the study.

Data Availability statement

The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

References

Alabdulaziz, M.S. (2023), "Escape rooms technology as a way of teaching mathematics to secondary school students", *Education and Information Technologies*, Vol. 28 No. 10, pp. 13459-13484, doi: 10.1007/s10639-023-11729-1.

Bipp, T., Wee, S., Walczok, M. and Hansal, L. (2024), "The relationship between game-related assessment and traditional measures of cognitive ability—a meta-analysis", *Journal of Intelligence*, Vol. 12 No. 12, 129, doi: 10.3390/jintelligence12120129.

Blanco, C., Almarza, M., Massana, J., Alonso, F. and Martinez, E. (2023), "Digital escape rooms as a way to increase academic performance and student's motivation", *Proceedings of the 16th annual International Conference of Education, Research and Innovation (ICERI 2023)*, Seville, 13-15 November 2023, Vol. 1, pp. 2125-2131.

Deterding, S., Dixon, D., Khaled, R. and Nacke, L. (2011), "From game design elements to gamefulness: defining 'gamification'", *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments (MindTrek'11)*, Tampere, Finland, 28-30 September 2011, Vol. 1, pp. 9-15.

Edo-Agustín, E., Bonaiuti, G. and Marras, A. (2025), "Exploring the potential of edu-breakout and classroom quizzes to gamify teaching at university", *Qwerty*, Vol. 2025, 96, doi: 10.30557/qw000096.

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H. and Wenderoth, M.P. (2014), "Active learning increases student performance in science, engineering, and mathematics", *Proceedings of the National Academy of Sciences (PNAS)*, Vol. 111 No. 23, pp. 8410-8415, doi: 10.1073/pnas.1319030111.

Gee, J.P. (2003), *What Video Games Have to Teach Us about Learning and Literacy*, Palgrave Macmillan, London.

Gee, J.P. (2005), "Are video games good for learning?", *Nordic Journal of Digital Literacy*, Vol. 1 No. 3, pp. 172-183.

Giner-Baixauli, A., Corbí, H. and Mayoral, O. (2025), "The history of Earth and life: escape room for high school students", *Education Sciences*, Vol. 15 No. 4, 485, doi: 10.3390/educsci15040485.

González-Pérez, L.I. and Ramírez-Montoya, M.S. (2022), "Components of Education 4.0 in 21st century skills frameworks: systematic review", *Sustainability*, Vol. 14 No. 3, 1493, doi: 10.3390/su14031493.

Grande-de-Prado, M., Abella-García, V., Baelo-Álvarez, R. and García-Martín, S. (2025), "Pre-service teacher perceptions of the educational escape rooms", *Journal of Technology and Science Education*, Vol. 15 No. 2, pp. 543-556, doi: 10.3926/jotse.3020.

Gueye, M.L. and Exposito, E. (2023), "Education 4.0: proposal of a model for autonomous management of learning processes", *Lecture Notes in Computer Science*, Vol. 13821, pp. 106-117.

Hacisalihoglu, G., Stephens, D., Johnson, L. and Edington, M. (2018), "The use of an active learning approach in a SCALE-UP learning space improves academic performance in undergraduate General Biology", *Public Library of Science ONE (PLoS ONE)*, Vol. 13 No. 5, e0197916, doi: 10.1371/journal.pone.0197916.

Johnson, D.W. and Johnson, R.T. (2018), "Cooperative learning: the foundation for active learning", in Brito, S.M. (Ed.), *Active Learning - beyond the Future*, IntechOpen, London.

Lansdell, B.J. and Kording, K.P. (2019), "Towards learning-to-learn", *Current Opinion in Behavioral Sciences*, Vol. 29, pp. 45-50, doi: 10.1016/j.cobeha.2019.04.005.

Liu, J. and Liu, X. (2025), "Enhancing active learning through gamification: developing research skills with an escape room experience", *Journal of Learning Development in Higher Education*, Vol. 37, 1721, doi: 10.47408/jldhe.vi37.1721.

Mañas-Rodríguez, M.Á., Alcaraz-Pardo, L., Pecino-Medina, V. and Limbert, C. (2016), "Validation of the Spanish version of Soane's ISA engagement scale", *Journal of Work and Organizational Psychology*, Vol. 32 No. 2, pp. 87-93, doi: 10.1016/j.rpto.2016.04.002.

Matsumoto-Royo, K., Ramírez-Montoya, M.S. and Conget, P. (2021), "Opportunities to develop lifelong learning tendencies in practice-based teacher education: getting ready for Education 4.0", *Future Internet*, Vol. 13 No. 11, 292, doi: 10.3390/fi13110292.

Michael, J. (2006), "Where's the evidence that active learning works?", *Advances in Physioilogy Education*, Vol. 30 No. 4, pp. 159-167, doi: 10.1152/advan.00053.2006.

Mukul, E. and Büyüközkan, G. (2023), "Digital transformation in education: a systematic review of Education 4.0", *Technological Forecasting and Social Change*, Vol. 194, 122664, doi: 10.1016/j.techfore.2023.122664.

Nicholson, S. (2015), "Peeking behind the locked door: a survey of escape room facilities", available at: https://scottnicholson.com/pubs/erfacwhite.pdf (accessed 15 March 2026).

Nurbavliyev, O., Kaymak, S. and Sydykov, B. (2022), "The effect of active learning method on students' academic success, motivation and attitude towards mathematics", *Journal of Language and Linguistic Studies*, Vol. 18 No. 2, pp. 701-713.

Nwachukwu, C. and Osa-Izeko, E. (2022), "Assessment of intellectual, social and affective (ISA) engagement of academics in Nigerian universities", *Nigerian Journal of Management Sciences*, Vol. 23 No. 2, pp. 20-28, available at: https://nigerianjournalofmanagementsciences.com/wp-content/uploads/2022/09/3

Oliveira, W. and Hamari, J. (2025), "Flow experience in gameful approaches: a systematic literature review, scientometric analysis, and research agenda", *International Journal of Human-Computer Interaction*, Vol. 41 No. 20, pp. 13113-13139, doi: 10.1080/10447318.2025.2470279.

Pacheco-Velazquez, E., Paragarino, V.R., Glasserman, L.D. and Arroyo, M.C. (2024), "Playing to learn: developing self-directed learning skills through serious games", *Journal of International Education in Business*, Vol. 17 No. 3, pp. 416-430, doi: 10.1108/jieb-08-2023-0054.

Patiño, A., Ramírez-Montoya, M.S. and Buenestado-Fernández, M. (2023), "Active learning and Education 4.0 for complex thinking training: analysis of two case studies in open education", *Smart Learning Environments*, Vol. 10 No. 1, 8, doi: 10.1186/s40561-023-00229-x.

Peláez-Sánchez, I.C., Velarde-Camaqui, D. and Glasserman-Morales, L.D. (2024), "The impact of large language models on higher education: exploring the connection between AI and Education 4.0", *Frontiers in Education*, Vol. 9, 1392091, doi: 10.3389/feduc.2024.1392091.

Polytechnic University of Valencia – Office of International Exchange Programmes (2020), "Spanish grading system in higher education", available at: https://www.upv.es/entidades/OPII/infoweb/pi/info/1147768normali.html (accessed 15 March 2026).

Rivera, E.S. and Garden, C.L.P. (2021), "Gamification for student engagement: a framework", *Journal of Further and Higher Education*, Vol. 45 No. 7, pp. 999-1012, doi: 10.1080/0309877x.2021.1875201.

Rivest, L. and Ebouele, S.E. (2020), "Sampling a two dimensional matrix", *Computational Statistics and Data Analysis*, Vol. 149, 106971, doi: 10.1016/j.csda.2020.106971.

Rogti, M. (2021), "The process of teaching in the light of modern pedagogy: EFL teachers pedagogical reflection", *Proceedings of the 5th International Academic Conference on Humanities and Social Sciences (IACHSS 2021)*, Brussels, Belgium, 18-20 June 2021.

Ryan, R.M. and Deci, E.L. (2000), "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being", *American Psychologist*, Vol. 55 No. 1, pp. 68-78, doi: 10.1037//0003-066x.55.1.68.

Sarin, J., Charan, G.S., Bharti, R., Sharma, S. and Kaur, E. (2025), "Innovative learning with escape rooms in simulations, technologies, and pedagogical impact in Healthcare education: an integrated review", *Annals of African Medicine*, Vol. 24 No. 3, pp. 540-548, doi: 10.4103/aam.aam_66_25.

Shidarta, I. (2019), "The intellectual, social, affective engagement scale (ISA engagement scale): a validation study", *Jurnal Computech and Bisnis*, Vol. 13 No. 1, pp. 50-57.

Sipone, S., Gutiérrez, A.R., Berodia, J.L.M. and Arce, M.R. (2024), "The use of escape room as an alternative teaching strategy for sustainable mobility", *Transportation Research Procedia*, Vol. 78, pp. 538-545, doi: 10.1016/j.trpro.2024.02.067.

Soane, E., Truss, C., Alfes, K., Shantz, A., Rees, C. and Gatenby, M. (2012), "Development and application of a new measure of employee engagement: the ISA Engagement Scale", *Human Resource Development International*, Vol. 15 No. 5, pp. 529-547, doi: 10.1080/13678868.2012.726542.

Soares-Collado, A. and Sánchez-Hernández, M.I. (2020), "The escape room in the human capital management: an application in personnel selection", *Administración Y Organizaciones*, Vol. 23 No. 44, pp. 100-124, doi: 10.24275/uam/xoc/dcsh/rayo/2020v23n44/soares.

Stieha, V., Earl, B., Hagens, H., Haynes, M., Ulappa, A., Bond, L. and Oxford, J.T. (2024), "An exploration of the relationship between active learning and student motivation in STEM: a mixed methods study", *AJP Advances in Physiology Education*, Vol. 48 No. 3, pp. 621-638, doi: 10.1152/advan.00247.2022.

Ugo, C.C., López-Pernas, S., Gordillo, A. and Tukiainen, M. (2025), "Escape rooms for learning programming: a systematic literature review", *Review of Education*, Vol. 13 No. 3, e70123, doi: 10.1002/rev3.70123.

Veldkamp, A., Van De Grint, L., Knippels, M.P. and Van Joolingen, W.R. (2020), "Escape education: a systematic review on escape rooms in education", *Educational Research Review*, Vol. 31, 100364, doi: 10.1016/j.edurev.2020.100364.

Yu, M., Guo, C., Fan, W. and Han, S. (2026), "Experiences and perceptions of escape room-based learning among nursing students: a systematic review and meta-synthesis", *Nurse Education Today*, Vol. 161, 107016, doi: 10.1016/j.nedt.2026.107016.

Corresponding author
Pedro Juan Roig can be contacted at: proig@umh.es