

Rapid Academic Levelling Through Artificial Intelligence Role Play Escape Room

Carlos Astengo¹ and Mario Caudillo²

¹Tecnologico de Monterrey, Nuevo León, Mexico

²Universidad Internacional de La Rioja (UNIR), Logroño, Spain

castengo@tec.mx

mario.caudillo@unir.net

Abstract: Gaps in previously taught probability and statistics concepts can present difficulties when engineering students enter courses that rely on this prerequisite knowledge. Educational escape rooms provide one possible mechanism for revisiting such concepts through contextualised problem-solving within a limited amount of classroom time. This exploratory study examines aggregate changes in assessment performance and response behaviour surrounding a single-session conversational artificial intelligence (AI) educational escape-room intervention with second-semester engineering students in Mexico. The complete cohort of 15 students enrolled in an Engineering in Interactive Software and Video Game Development programme participated voluntarily. Participants were aged 18–22 years; two identified as women, twelve as men, and one as non-binary. The one-hour classroom session consisted of a 10-minute pre-test, a 5-minute introduction, a 35-minute individually completed AI-mediated escape room, and a 10-minute post-test. The intervention, *Probabilidades Sangrientas: El Castillo de Strahd*, was implemented as a customised GPT-based conversational experience containing 20 open-response probability and statistics challenges embedded within a gothic escape-room narrative. Fifteen pre-test records and fourteen post-test records were available for analysis; individual responses could not be matched. Consequently, analysis was restricted to descriptive aggregate comparisons. The proportion of attempted items increased from 40.0% to 68.6%, while correct responses as a proportion of all possible items increased from 19.3% to 55.7%. Among attempted items, correctness increased from 48.3% to 81.3%. These differences indicate improved immediate aggregate performance and greater willingness to attempt assessment items after the session. However, because the study lacked a control group, matched individual observations, and a delayed assessment, the findings cannot establish individual learning gains or attribute the observed changes specifically to the escape-room or AI components. The study instead provides an exploratory account of how a conversational AI escape room can be incorporated into a time-constrained engineering classroom for prerequisite-knowledge reinforcement.

Keywords: Higher education, Educational innovation, Game-based learning, Digital escape room, Probability and statistics, Conversational artificial intelligence

1. Introduction

Engineering curricula commonly assume that students entering mathematically intensive subjects retain concepts acquired in preceding courses. In practice, however, prerequisite knowledge may be uneven when students progress from one semester to another. This creates a practical instructional challenge: teachers may need to reactivate prior knowledge without devoting substantial portions of a new course to reteaching material formally addressed elsewhere in the curriculum.

Game-based learning provides one possible approach for incorporating such reinforcement into active classroom experiences. Previous studies have associated educational games with knowledge acquisition and motivation in computer science and software engineering education (Gordillo et al., 2022; López-Fernández et al., 2023). Educational escape rooms represent a more specific form of game-based activity in which disciplinary problems are embedded within a narrative progression and associated with advancement towards an explicit goal.

Systematic reviews have identified increasing educational use of both physical and digital escape rooms across disciplinary settings (Fotaris & Mastoras, 2019; Makri et al., 2021). Mathematics-specific research has similarly examined escape-room approaches as active methodologies for mathematical learning (Fuentes-Cabrera et al., 2020), while Bakar et al. (2023) reported improvements in mathematics performance alongside positive motivational and collaborative perceptions following a digital escape-game intervention. Within engineering education, escape-room approaches have also been implemented in areas including software engineering and engineering ethics (Gordillo & López-Fernández, 2024; Monteiro & Sousa, 2024).

Importantly, the strength of conclusions that can be drawn from such interventions depends on research design. For example, Gordillo and López-Fernández (2024) compared an educational escape room against traditional instruction using a randomised controlled design. Designs of this kind permit considerably stronger claims regarding comparative effectiveness than uncontrolled classroom implementations. Guidance

concerning serious-game evaluation similarly emphasises the relationship between evaluation design and the conclusions that resulting data can support (Baalsrud Hauge et al., 2014).

The present study addresses a different practical context. Probability and statistics had been formally introduced to the participating students during their preceding Mathematics Discrete and Statistics course. At the beginning of their subsequent Linear Algebra and Geometry course, these topics were considered important prerequisite knowledge, but the course schedule did not permit their complete reteaching. A one-hour session was therefore allocated to diagnosis and reinforcement.

The term rapid academic levelling is used here descriptively to refer to this instructional constraint: the intervention was intentionally incorporated into a single regular class session rather than implemented as a separate remedial programme. It is not proposed as an established duration-based category of educational intervention.

The escape-room component differs from conventional digital escape rooms in that interaction occurred through a customised conversational GPT system. Rather than evaluating the effectiveness of this system, the present exploratory study asks:

RQ1. What aggregate differences in assessment performance and willingness to attempt items are observable between the pre-test and immediate post-test surrounding a single-session conversational AI educational escape-room intervention?

A secondary contribution of the study is to document the design and implementation of the AI-mediated escape room sufficiently to clarify the respective roles of its narrative, mathematical challenges, conversational AI, feedback, scoring and progression mechanisms.

2. Methods

2.1 Study Design and Participants

The study employed an exploratory one-group pre-test/post-test classroom design without a control or comparison group. It was conducted during the second meeting of a Linear Algebra and Geometry course in an Engineering in Interactive Software and Video Game Development programme at a private university in the State of Mexico.

The participants constituted the complete cohort of students enrolled in the second-semester course. All 15 students were invited to participate and all voluntarily agreed. Eligibility therefore corresponded to membership of the class and prior completion of the preceding Mathematics Discrete and Statistics course, in which probability and statistics had been introduced. No additional inclusion or exclusion criteria were applied.

Participants were between 18 and 22 years of age. Two identified as women, twelve as men, and one as non-binary.

All 15 participants generated a pre-test record. At the end of the activity, all students showed the instructor a completion screen for the post-test; however, when the Google Forms dataset was subsequently accessed, only 14 post-test records were present. No participant was observed to abandon the activity, and the reason for the missing record could not be established. The analysis therefore uses 15 available pre-test records and 14 available post-test records rather than treating the difference as participant attrition.

Individual identifiers were not collected in the assessment forms, preventing reliable matching between pre-test and post-test responses.

2.2 Classroom Procedure

The complete intervention occupied one 60-minute period at the beginning of the class:

Students first completed the pre-test individually with 10-minute time limit. During the five-minute introduction, the escape-room objective and interaction procedure were explained. Students then interacted individually with the system using their own mobile phones for 35 minutes.

Participants were informed that they could ask the conversational character for guidance or explanations when they did not know how to approach a challenge. The activity was followed immediately by the post-test with the same 10-minute time limit.

2.3 Escape-room Design

The intervention was titled *Probabilidades Sangrientas: El Castillo de Strahd*. It had originally been designed by Carlos Astengo for a different instructional context and had previously undergone iterative testing by its designer as well as classroom pilot use with students at his institution. These earlier uses were part of instructional development rather than a formal validation study; consequently, no psychometric or experimental validation metrics are available from that phase.

The experience situates the learner inside the fictional castle of the vampire Count Strahd. Mathematical problems function as obstacles to progression through the castle rather than as disconnected exercises. The overarching objective is to overcome 20 probability and statistics challenges and escape.

The system specifies four narrative final outcomes based on cumulative performance as denoted in Figure 1.

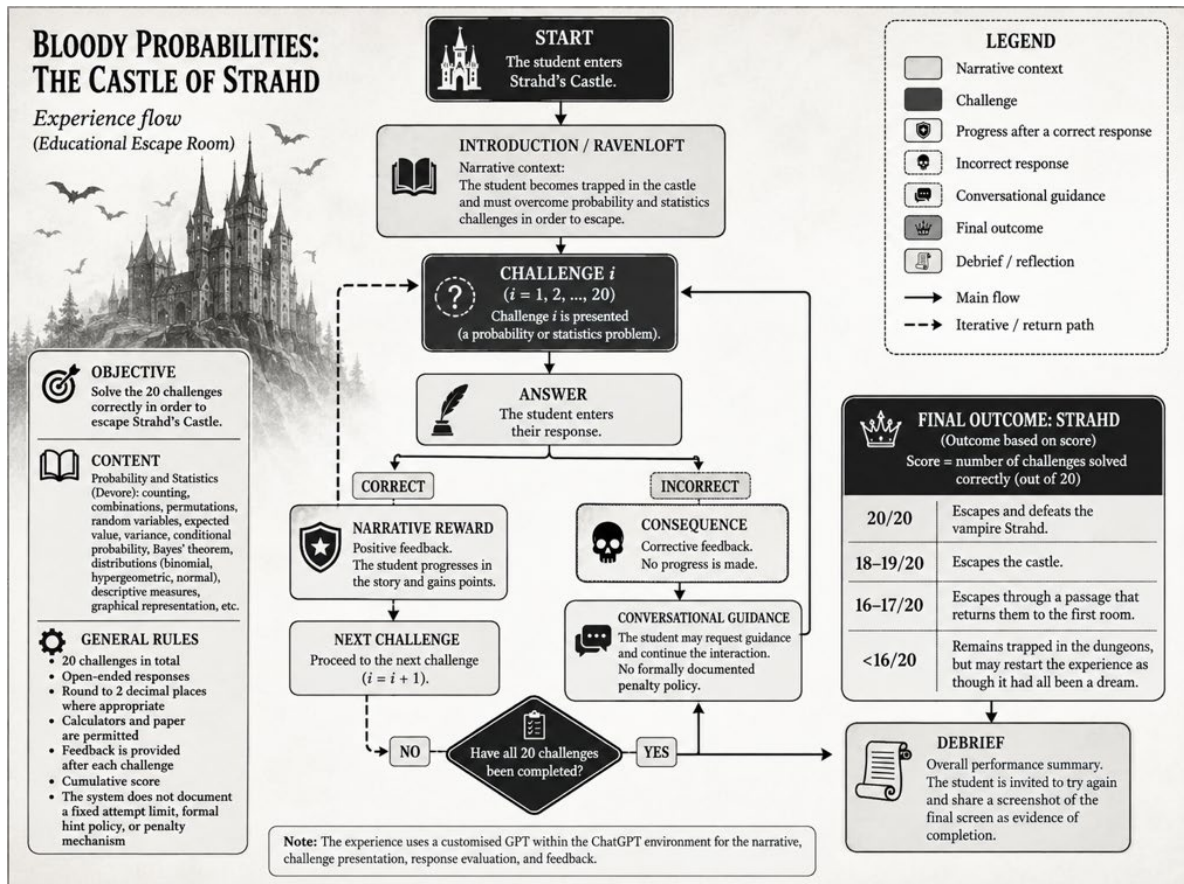


Figure 1: Conceptual interaction flow of the AI-mediated escape room

The maximum success condition therefore required 20 correct challenges. At completion, the system generated an overall evaluation and invited the learner to attempt the experience again. Students were instructed to capture the final screen as evidence that they had completed the activity.

The configured mathematical content included counting techniques, the multiplicative principle, combinations, permutations, random variables, expected value, variance, conditional probability, Bayes' theorem, binomial distribution, hypergeometric distribution, normal distribution, variable types, measures of central tendency, measures of dispersion and graphical representation. The available system configuration identifies Jay L. Devore's *Probability and Statistics for Engineering and the Sciences* as its disciplinary basis; however, the exact edition used during development was not preserved in the available study documentation.

The experience contained exactly 20 open-response challenges. Responses were not multiple choice. Numerical answers were expected to be expressed to two decimal places when appropriate, and the use of paper and calculators was permitted. Feedback was provided following responses and the system maintained a cumulative score.

The surviving configuration does not document a fixed number of attempts, an attempt limit, a formal penalty mechanism, automatic hint frequency or a predetermined rule governing retries. Although students could conversationally request assistance, these behaviours are therefore not represented as formally standardised game mechanics. A simplified interaction architecture is also shown in Figure 1.

When additional guidance was requested, interaction could continue conversationally before progression. Because no standardised hint/retry protocol was recorded, this branch should be interpreted as conversational support rather than a fixed rule-based hint mechanic.

2.4 Role of the Conversational AI

The intervention was implemented as a customised GPT-based educational experience within the ChatGPT environment. The available system is identified as *Probabilidades Sangrientas: El Castillo de Strahd*. The historical documentation available for the study does not preserve the exact underlying model version used during the classroom intervention. Consequently, no specific GPT model designation is reported.

The system configuration assigned the conversational AI several functions:

- presenting mathematical challenges within the escape-room narrative;
- interpreting open-ended student responses;
- determining whether responses were correct;
- producing explanatory feedback;
- maintaining the narrative context;
- maintaining the cumulative score; and
- generating the corresponding final narrative outcome.

The narrative framework, total number of challenges, targeted mathematical content, scoring structure and outcome thresholds were predefined in the custom configuration.

The extent to which individual numerical problem instances were identical across participants cannot be reconstructed from the retained data. During iterative development, numerical quantities could vary between generated interactions. However, interaction logs from the experimental session were not collected, and it therefore cannot be confirmed whether all participants encountered identical numerical instances or how much generative variability occurred during the classroom implementation. Likewise, the available configuration does not provide evidence of automatic difficulty adaptation.

During development, the system designer repeatedly tested the system and manually checked the mathematical validity of responses, explanations and assistance generated by the GPT. Because numerical values could vary during interaction, exhaustive validation of every possible generated instance was not feasible. No formal automated mathematical verification procedure or comprehensive validation log was implemented.

2.5 Pre-test and Post-test Instruments

The pre-test and post-test each consisted of ten open-ended items intended to sample prerequisite probability and statistics knowledge. Both were designed for completion within ten minutes.

The two forms were structurally aligned by mathematical topic and required analogous operations while changing numerical values or contextual framing.

Table 1: Pre-test and Post-test constructs.

Item	Pre-test construct	Post-test construct
1	Multiplicative counting principle	Multiplicative counting principle
2	Permutations	Permutations
3	Expected value	Expected value
4	Conceptual interpretation of variance	Conceptual interpretation of dispersion
5	Binomial probability	Binomial probability
6	Hypergeometric probability	Hypergeometric probability

Item	Pre-test construct	Post-test construct
7	Standardisation / Z score	Standardisation / Z score
8	Conditional probability	Conditional probability
9	Bayes' theorem	Bayes' theorem
10	Arithmetic mean	Arithmetic mean

For example, the first pre-test item asked students to determine the number of ways to select one probability book and one linear-algebra book from sets of four and three books respectively. Its post-test counterpart required determination of the number of outfits possible from five shirts and two pairs of trousers. Both therefore required application of the multiplicative counting principle.

Similarly, permutation, expected-value, binomial, hypergeometric, standardisation, conditional probability, Bayesian probability and descriptive-statistics items were paired by topic while changing the quantities and scenario.

This alignment documents structural and content similarity between the forms but should not be interpreted as evidence of psychometric equivalence. No formal validation, item-response analysis or psychometric equivalence test was conducted.

2.6 Scoring

Responses were classified into three categories: correct, incorrect or unanswered. No partial credit was awarded.

A predefined solution was available for each item. Two evaluators participated in scoring. For conceptually phrased responses, correctness was established through consensus between the two evaluators. Inter-rater reliability was not calculated.

Students had been instructed not to guess and to leave an item unanswered when they did not know how to solve it. Consequently, unanswered items were retained as a distinct response category.

2.7 Data Analysis

Because individual students could not be matched between the pre-test and post-test datasets, and because one fewer post-test record was available, no paired inferential analysis was conducted. Statistical significance tests and effect-size estimates were therefore excluded from the revised analysis.

The analysis was intentionally descriptive and focused on four indicators.

The first indicator, the attempt rate was calculated as the number of attempted responses divided by the total number of possible responses. The total number of possible responses corresponded to the number of available student records multiplied by the ten assessment items.

As second indicator, the correct-response rate across all possible items was calculated as the number of correct responses divided by the total number of possible responses. This measure reflects overall assessment performance while retaining unanswered items in the denominator.

The third, correctness among attempted responses was calculated as the number of correct responses divided only by the number of responses actually attempted. This distinction was introduced to avoid treating unanswered items as incorrect responses when describing the accuracy of submitted answers.

And as the fourth indicator, the mean number of correct responses per available record was calculated by dividing the total number of correct responses by the number of student records available at each measurement point.

Changes between the pre-test and post-test were reported descriptively as percentage-point differences where appropriate. These differences represent comparisons between aggregate datasets and must not be interpreted as individual learning gains.

3. Results

Fifteen pre-test records and fourteen post-test records were available for analysis. Although all fifteen students showed evidence of having completed the post-test at the end of the classroom session, only fourteen records were subsequently present in the Google Forms dataset. The reason for the missing record could not be established. Because the assessments were anonymous, individual pre-test and post-test responses could not be matched.

In the pre-test, 150 responses were possible across the fifteen participants and ten assessment items. Students attempted 60 of these responses, corresponding to an attempt rate of 40.0%. In the post-test, 140 responses were possible across the fourteen available records, of which 96 were attempted, corresponding to an attempt rate of 68.6%. The proportion of attempted items was therefore 28.6 percentage points higher in the post-test. As no validated measure of engagement or confidence was administered, this difference is interpreted only as a change in response behaviour and willingness to attempt assessment items, rather than as direct evidence of increased engagement or confidence.

Performance also differed between the two measurements. In the pre-test, 29 of the 150 possible responses were correct, corresponding to a correct-response rate of 19.3% across all possible items. In the post-test, 78 of the 140 possible responses were correct, corresponding to 55.7%. This represents a descriptive difference of 36.4 percentage points between the two datasets.

Because students had been instructed not to guess and could leave questions unanswered when they did not know how to solve them, correctness was also examined exclusively among attempted responses. In the pre-test, 29 of the 60 attempted responses were correct, equivalent to 48.3%. In the post-test, 78 of the 96 attempted responses were correct, equivalent to 81.3%. Correctness among attempted responses was therefore 32.9 percentage points higher in the post-test.

The mean number of correct responses per available record also differed between measurements. The pre-test produced 29 correct responses across 15 records, resulting in a mean of 1.93 correct answers per record, whereas the post-test produced 78 correct responses across 14 records, resulting in a mean of 5.57. The descriptive difference between these aggregate means was therefore 3.64 additional correct responses per available record. This value should not be interpreted as an average individual learning gain because the anonymous data did not permit individual pre-test and post-test scores to be paired.

Item-level results showed that differences were not restricted to a single topic. The binomial and hypergeometric probability items received no correct responses in the pre-test, whereas each received six correct responses in the post-test. The permutation item increased from six correct responses in the pre-test to twelve in the post-test. The descriptive-statistics item, which already showed comparatively strong performance before the intervention, increased from ten to eleven correct responses. These patterns provide additional descriptive information about the topics in which performance differed following the activity; however, they cannot establish within-student improvement or determine whether the observed differences resulted specifically from the escape-room intervention.

Table 2 summarises the aggregate descriptive results. Overall, the available post-test dataset contained a higher proportion of attempted responses, a higher proportion of correct responses across all possible items, and greater correctness among attempted responses than the pre-test dataset. Given the absence of matched observations and a comparison group, these results are presented strictly as descriptive differences between the two measurement points rather than as evidence of causal intervention effects.

Table 2: Aggregate descriptive comparison of pre-test and post-test results

Measure	Pre-test	Post-test	Descriptive difference
Available records	15	14	—
Possible responses	150	140	—
Attempted responses	60	96	—
Attempt rate	40.0%	68.6%	+28.6 percentage points
Correct responses	29	78	—

Measure	Pre-test	Post-test	Descriptive difference
Correct responses as a proportion of all possible items	19.3%	55.7%	+36.4 percentage points
Correct responses as a proportion of attempted items	48.3%	81.3%	+32.9 percentage points
Mean correct responses per available record	1.93	5.57	+3.64

Note: Pre-test and post-test datasets could not be matched at the individual level. Differences therefore represent aggregate descriptive comparisons and should not be interpreted as individual learning gains or causal effects of the intervention.

4. Discussion

The study identified substantial descriptive differences between the available pre-test and post-test datasets. Students attempted a larger proportion of questions after the intervention, and a larger proportion of both possible and attempted responses were correct.

The increase in attempted items is deliberately interpreted as response behaviour or willingness to attempt assessment items, rather than as a direct measure of engagement or confidence. Attempting an item cannot independently establish either psychological construct. No validated engagement, motivation or confidence instrument was administered in the present study.

This distinction is particularly important because the original classroom instruction asked participants not to guess when they did not know an answer. The simultaneous increase in willingness to respond and in correctness among attempted responses is therefore consistent with improved immediate ability to address the assessment tasks. Nevertheless, these measures do not reveal why individual students changed their response behaviour.

The item-level patterns provide additional descriptive evidence. In particular, binomial and hypergeometric probability questions moved from zero correct responses in the pre-test to six each in the post-test. These topics were also represented within the intervention's configured content. However, the design does not permit the improvement to be attributed specifically to the corresponding escape-room challenges.

The observed differences are compatible with several explanations. Students received additional exposure to probability and statistics during the 35-minute escape room, solved related problems, received feedback and then encountered structurally similar material in the post-test. Consequently, immediate improvement could reflect content exposure, retrieval practice, feedback, repeated testing, the narrative game environment, conversational assistance from the GPT, or combinations of these factors.

Without a comparison condition, the study cannot isolate the contribution of game mechanics from the contribution of additional mathematical practice. Similarly, it cannot determine whether the conversational AI provided an advantage over equivalent teacher-led practice, a non-AI digital escape room or conventional exercises.

This distinction separates the present exploratory study from controlled studies such as Gordillo and López-Fernández (2024), where experimental and comparison conditions permit stronger conclusions regarding relative instructional effectiveness.

The practical contribution is therefore narrower but still relevant. The intervention demonstrates one way in which prerequisite concepts can be revisited within a single regular class period rather than through a separate remedial unit. The complete diagnostic-and-reinforcement sequence occupied 60 minutes, of which 35 minutes corresponded to the escape-room interaction. In this paper, *rapid* therefore describes a deliberate curricular constraint rather than an experimentally established category of effective intervention duration.

The conversational AI architecture also introduces design possibilities that differ from more fixed digital escape rooms. Open-response interaction enables the system to interpret free-form mathematical answers, maintain a narrative persona and provide conversational feedback. At the same time, these capabilities create reproducibility and validation challenges. Generative variability can mean that learners do not necessarily experience identical interaction sequences, and automated mathematical evaluation requires explicit validation procedures.

5. Limitations and Future Research

Several limitations constrain interpretation of the findings.

The sample consisted of only 15 students from a single engineering cohort. Although this represented the complete population enrolled in the class, the sample is too small and context-specific to support broad generalisation.

There was no control or comparison group. The observed pre-to-post differences therefore cannot be attributed specifically to the escape room, its game mechanics or the conversational AI.

Assessment responses were anonymous and could not be paired. Although all students appeared to complete the post-test, one post-test record was absent from the retained dataset for an unknown reason. The study consequently cannot calculate individual learning gains or determine whether improvement was concentrated among particular participants.

The pre-test and post-test were structurally aligned by topic but were not psychometrically validated as equivalent forms. Their ten-minute format was designed primarily for classroom feasibility.

The study measured immediate post-intervention performance only. No delayed post-test was conducted, preventing conclusions about retention.

No validated measures of engagement, motivation or confidence were collected. Attempt behaviour should therefore not be interpreted as a substitute for these constructs.

Complete interaction logs between students and the GPT were not collected. As a result, the study cannot reconstruct the exact challenges, numerical variations, hints, feedback sequences or number of conversational exchanges experienced by each participant.

The precise historical GPT model version used during the intervention was not preserved. Moreover, although the designer iteratively checked mathematical responses during development, no exhaustive validation system was available for all possible dynamically varying numerical instances.

Future implementations should therefore assign anonymous participant codes to permit paired analysis while preserving anonymity; include a comparison condition; incorporate a delayed retention test; retain privacy-preserving interaction logs; document the exact deployed model and system configuration; formalise hint, retry and feedback rules; and implement systematic mathematical validation of generated content.

A future study could also compare at least three conditions: conventional practice exercises, a fixed digital escape room and a conversational AI escape room. Such a design would make it possible to investigate whether observed differences arise from additional practice, game design or the conversational AI layer.

6. Conclusion

This exploratory classroom study examined aggregate differences surrounding a one-hour probability and statistics levelling session incorporating a 35-minute conversational AI educational escape room. The available data show clear descriptive changes. The proportion of attempted assessment items increased from 40.0% to 68.6%, correct responses as a proportion of all possible items increased from 19.3% to 55.7%, and correctness among attempted items increased from 48.3% to 81.3%. These findings demonstrate that stronger immediate aggregate assessment performance followed the intervention, but they do not establish that the escape room caused the improvement. Additional exposure, practice, feedback and repeated assessment are plausible alternative explanations. The main contribution of the study is therefore not evidence of intervention effectiveness, but an exploratory demonstration and transparent documentation of how a conversational AI educational escape room can be integrated into a time-constrained engineering classroom to revisit prerequisite mathematical knowledge. Further controlled research with matched observations, validated instruments, retained interaction logs and delayed assessment is required before claims can be made concerning learning effectiveness, retention or the distinctive contribution of conversational AI.

Acknowledgements

We want to acknowledge the technical and financial support of the Writing Lab, Institute for the Future of Education, Tecnológico de Monterrey, Mexico, in the production of this work. Also, we want to acknowledge the technical and financial support of Universidad Internacional de La Rioja (UNIR).

Ethics Declaration: Under the institutional procedures applicable to this classroom educational activity, formal Ethics Committee review was not required. Before the intervention, students were informed that participation in the research component was voluntary and that anonymised assessment results could be used for research and publication. All participants were at least 18 years old and provided consent.

Participation did not affect course grades. Assessment forms did not collect student names or identifiable email addresses. Data were initially collected through Google Forms and stored in a restricted-access Google Drive file. The exported research dataset was subsequently retained by the authors in anonymised form on their computers.

The authors report no competing interests.

AI Declaration: Conversational artificial intelligence constituted part of the educational intervention evaluated in this study, as described in the Methods section.

For preparation and revision of the manuscript, the authors used OpenAI's ChatGPT to support structural reorganisation, language revision and drafting based on study information supplied and verified by the authors. Scite was used to support searches for relevant academic literature. The authors reviewed and verified the factual descriptions, calculations, interpretations and references and retain responsibility for the final manuscript.

References

- Bakar, Dk. N. N. P. A., Shahrill, M., & Zakariya, Y. F. (2023). Digital Escape Game and Students' Learning Outcomes in Mathematics: Experience From Brunei. *Sage Open*, 13(4). <https://doi.org/10.1177/21582440231216838>
- Baalsrud Hauge, J., Boyle, E., Mayer, I., Nadolski, R., Riedel, J. C. K. H., Moreno-Ger, P., Bellotti, F., Lim, T., & Ritchie, J. (2014). Study design and data gathering guide for serious games' evaluation. In T. M. Connolly, T. Hainey, E. Boyle, G. Baxter, & P. Moreno-Ger (Eds.), *Psychology, pedagogy, and assessment in serious games* (pp. 394–419). IGI Global. <https://doi.org/10.4018/978-1-4666-4773-2.ch018>
- Carrillo-Nieves, D., Clarke-Crespo, E., Cervantes-Avilés, P., Cuevas-Cancino, M. de la O., & Vanoye-García, A. Y. (2024). Designing learning experiences on climate change for undergraduate students of different majors. *Frontiers in Education*, 9. <https://doi.org/10.3389/educ.2024.1284593>
- Cocca, A., Verdugo, F. D. E., Cuenca, L. T. R., & Cocca, M. (2020). Effect of a Game-Based Physical Education Program on Physical Fitness and Mental Health in Elementary School Children. *International Journal of Environmental Research and Public Health*, 17(13), 4883. <https://doi.org/10.3390/ijerph17134883>
- Fotaris, P., & Mastoras, T. (2019). Escape rooms for learning: A systematic review. In *Proceedings of the European Conference on Games-Based Learning* (pp. 235–243). https://www.researchgate.net/publication/336374954_Escape_Rooms_for_Learning_A_Systematic_Review
- Fuentes-Cabrera, A., Parra-González, M. E., López-Belmonte, J., & Segura-Robles, A. (2020). Learning mathematics with emerging methodologies—The escape room as a case study. *Mathematics*, 8(9), 1586. <https://doi.org/10.3390/math8091586>
- García, R. S. B., Rejas-Borjas, L. G., Cruzata-Martínez, A., & Mancisidor, M. C. S. (2022). The Use of Augmented Reality in Latin-American Engineering Education: A Scoping Review. *Eurasia Journal of Mathematics Science and Technology Education*, 18(1), em2064. <https://doi.org/10.29333/ejmste/11485>
- González-de la Torre, H., Hernández-De Luis, M.-N., Mies-Padilla, S., Camacho-Bejarano, R., Verdú-Soriano, J., & Rodríguez-Suárez, C.-A. (2024). Effectiveness of “escape room” educational technology in nurses' education: A systematic review. *Nursing Reports*, 14(2), 1193–1211. <https://doi.org/10.3390/nursrep14020091>
- Gordillo, A., López-Fernández, D., & Tovar, E. (2022). Comparing the Effectiveness of Video-Based Learning and Game-Based Learning Using Teacher-Authored Video Games for Online Software Engineering Education. *IEEE Transactions on Education*, 65(4), 524–532. <https://doi.org/10.1109/te.2022.3142688>
- Gordillo, A., & López-Fernández, D. (2024). Are Educational Escape Rooms More Effective Than Traditional Lectures for Teaching Software Engineering? A Randomized Controlled Trial. *IEEE Transactions on Education*, 67(5), 660–668. <https://doi.org/10.1109/te.2024.3403913>
- López-Fernández, D., Gordillo, A., Pérez, J., & Tovar, E. (2023). Learning and Motivational Impact of Game-Based Learning: Comparing Face-to-Face and Online Formats on Computer Science Education. *IEEE Transactions on Education*, 66(4), 360–368. <https://doi.org/10.1109/te.2023.3241099>
- Makri, A., Vlachopoulos, D., & Martina, R. A. (2021). Digital escape rooms as innovative pedagogical tools in education: A systematic literature review. *Sustainability*, 13(8), 4587. <https://doi.org/10.3390/su13084587>
- Monteiro, F., & Sousa, A. (2023). Pedagogical innovation to captivate students to ethics education in engineering. *Journal of Applied Research in Higher Education*, 16(4), 1029–1041. <https://doi.org/10.1108/jarhe-02-2023-0056>