Paper ID #44045

Use of Game-Based Learning with ChatGPT to Improve Mathematical Modeling Competences in First-Year Engineering Students

Dr. Gibrán Sayeg-Sánchez, Tecnologico de Monterrey

Dr. Gibrán Sayeg-Sánchez is professor – consultant in the Science Department in Tecnologico de Monterrey, Puebla campus. He studied a PhD in Financial Science in EGADE Business School (2016), a MSc in Industrial Engineering in Tecnologico de Monterrey (2011), and a BEng in Industrial and Systems Engineering in Tecnologico de Monterrey (2006). Dr. Sayeg-Sánchez has more than 11 years of experience in teaching statistics, mathematics, and operations research; and more than 13 years of experience in Operational Excellence consulting. His current research interests are focused in educational innovation and educational technologies.

Prof. Miguel X. Rodriguez-Paz, Tecnologico de Monterrey

Prof. RodrÃguez-Paz got his B.Sc. In Civil Engineering from Tecnologico de Oaxaca in 1993. He studied a M.Sc. In Structural Engineering at Tecnologico de Monterrey and got his Ph.D. from the University of Wales at Swansea in 2003 where he did research on

Use of Game – Based Learning with ChatGPT to Improve Mathematical Modelling Competences in First Year Engineering Students

Keywords: Higher Education, ChatGPT, Game – Based Learning, Educational Innovation

Introduction

The educational model in Tecnologico de Monterrey finds its foundation in Competence Based Learning [1], which requires that our students must present evidence of learning beyond the merely theoretical concepts of their courses. Competence based educational models are centered in developing three dimensions of students: theoretical frameworks, application skills, and attitudes and values [2], as depicted by Tecnologico de Monterrey's definition of competence shown in Figure 1. This combination of dimensions makes necessary a new form of delivering contents to students and evaluating the correct development of competences.

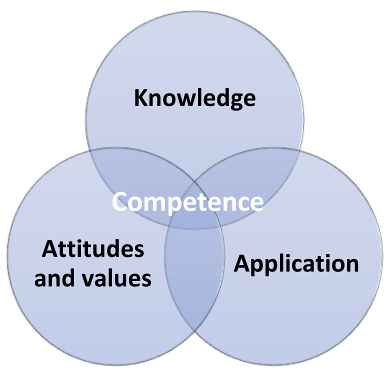

Figure 1. Definition of competence in Tecnologico de Monterrey.

A coherent integration of these three aspects allows students to demonstrate consistent behaviors when performing tasks or assessments. This characteristic of students is desirable, since consistent and repeatable behaviors towards educational activities imply the adequate development of a competence and facilitate the acquisition of greater domain levels of such competences [3]. Moreover, literature show that use of technology improves engagement of students and allows a better development of competences [4], hence, developing technological – based solutions for student's learning is important to motivate competence development.

This evidence-based paper explores the use of Artificial Intelligence (AI), within a context of Game-based learning, to develop mathematical competences in First Year Engineering students. The research includes the used methodology for prompt engineering and the obtained results of students' performance and perception of the methodology.

Literature Review

Game-Based learning is a methodology used in education that employs games as teaching tools [5]. Its use has been related to pedagogical benefits, such as the development of critical thinking skills [6], development of team building skills, fostering knowledge acquisition [7], and increasing motivation and engagement of students [8]

With the introduction of technology into the educational processes, methodologies such as Game Based Learning are positively affected, allowing the teachers to create more dynamic, realistic, and engaging activities [9]. Moreover, several authors have found that use of technology in education that allows involving the student in real or simulated situations increases their engagement and helps to develop competences and performance [2], [7], [10], [11], [12].

In recent years, Artificial Intelligence (AI) has been widely incorporated in education, which provides a broad range of opportunities to create innovative learning activities [13]. Studies have found that teachers often use AI for creating personalized learning experiences, providing quick feedback to students, and creating unique learning environments [14]. One of the technologies mostly used are the Generative Conversational AIs, such as ChatGPT, which presents both, benefits and concerns, towards its use in education.

Literature shows that ChatGPT 3.5, which provides free access to users, performs well under contexts of business, economics, and programming topics, however, it has several flaws when related to mathematics, which makes imperative a cautious use of the AI to develop competences in the students [15]. However, several efforts have been made to use the technology as an enhancer of mathematical skills such as mathematical reasoning and problem solving [16], [17].

One of the main implications in a successful implementation of ChatGPT activities in mathematical contexts is in generating the correct prompt, to assure that the technology will respond as expected by the teacher. Prompt engineering can be described as a combination of AI, linguistics, and UX [18]. One of the possible frameworks to craft efficient prompts is CLEAR, a 5 components model depicted in Table 1, that stands for Concise, Logical, Explicit, Adaptive, and Reflective [19].

Table 1. CLEAR framework for prompt engineering

Model	Component	Description
C	Concise	Prompts must be short and have clarity on what they state
L	Logical	Prompts must be structured and coherent
E	Explicit	Prompts must clearly specify inputs and outputs
A	Adaptive	Generating the prompt must be a flexible process that allows for modifications as it is being created
R	Reflective	Prompts must be continuously evaluated, and changes made to improve them through time

Methodology

As an important part of mathematical modelling, Tecnologico de Monterrey's curricula imply the solution of systems of differential equations. One of the specific models learned in the course is an epidemiologic model used to predict the behavior of a disease. This system of differential equations is known as the SIR model, which stands for: Susceptible people to contract the disease, Infected people with the disease, and Recovered people who cannot contract the disease again [20]. Therefore, we used the CLEAR framework to generate a prompt in ChatGPT that introduced Game-Based Learning with a simulated problem designed to develop mathematical modelling skills in First Year engineering students. The context of the game was the survival of a world pandemic by analyzing the behavior of a fictitious disease while interpreting the rates of change of the SIR model.

The prompt engineering process had six iterations, as shown in Table 2, in which the designed prompts were concise, were intended to provide a logical structure, and asked for explicit results. Each iteration allowed us to include more detail in the prompt and to redirect the objective of the game towards the development of mathematical modelling skills.

Table 2. Iterations of prompt engineering process

Iteration	**Improvement**	**Impact on game**
1	Base prompt	Provided basic game dynamics focused on SIR model
2	Included specific rates of growth for the SIR model	Provided a better framework for understanding the game and more detail in the game
3	Possible scenarios after each decision are generated as part of the game	Provided a more realistic experience when depicting the pandemic event
4	Random catastrophic events were added to the game dynamics	Increased the difficulty of the game, providing opportunities for modifying the player's strategy
5	Specific actions to counteract the pandemics effect from which the player may select	Provided a more structured game that guided the player throughout the desired challenges
6	Specific limits and initial values were provided for the game to start the iterations	Provided better control for the teacher to direct the activity towards intended learning purposes

The final prompt had a five steps logical structure, as shown in Figure 2, which contained: an *introduction*, where the context of the game is explained; a *gameplay dynamics* section, where the rules of the game and the main story is established; *actions to take*, where the participation of the student is defined and the range of possible solutions is established; a *game end* section, where conditions for the game to finish are defined; and a *scoring and Feedback* section, where we specify what kind of feedback must be given to the students.

Figure 2. Prompt logical structure

This final prompt for the designed activity is provided below, with each of the logical steps indicated in brackets, but not intended to include in the prompt. We encourage readers to try it on ChatGPT to experience the game and to introduce variations for being applied in other courses:

[Introduction] Act as a role-playing game. I will take on the role of a famous epidemiologist tasked with managing a global epidemic. To play my role, I must analyze and interpret results from the SIR model.

[Gameplay dynamics] You will create an initial situation of the pandemic and tell me the data for (S) susceptible, (I) infected, (R) recovered, and the trend of the variables (S), (I), and (R). In the initial situation, the population will not exceed 100,000 inhabitants, and the initial number of infected must be between 10 and 20% of the population.

You will ask me how to interpret the results of the SIR model and what action to take based on that interpretation. On each occasion, you will give me 3 possible actions to choose from. The chosen action will result in an increase or decrease in the number of infected. After that, you will update the SIR model, provide me with a scenario according to the decision made, and give me 3 new actions to choose from.

The growth of the infected occurs exponentially, with a high rate of relative growth. Random catastrophic events will be incorporated, such as mass travel of infected individuals, infections at sports events, defective vaccines, among others.

[Actions to take] Examples of actions to take include control measures, prevention measures, vaccination schemes, pandemic declaration, end of the pandemic, and economic impacts.

[Game end] The game ends when the pandemic is controlled or when the disease affects the entire population.

[Scoring and Feedback] Upon completion of the game, provide me with a score between 0 and 100 on how appropriate my decisions were for controlling the pandemic. Include suggestions on how I can better interpret the SIR model based on the decisions I made earlier.

For the implementation of the Game-Based activity, we followed a quasi – experimental study with a pre – test post – test analysis. The experiment involved 34 students that were enrolled in a Mathematical Modelling course during 2023 August – September period. Students were randomly divided in 2 groups: a control group of 8 students in which the modelling competence was evaluated by exposure to traditional classes but without access to Artificial Intelligence; and an experimental group with 26 students that, in addition to the same training of the control group, were exposed to the designed game.

During the experiment, it was of our interest to analyze the performance of students with respect to the Mathematical Modelling Competence. We focused in three skills of the competence: Understanding the meaning of rates of change, Interpretation of graphs, and decision making according to a mathematical model. Each skill was measured using an argumentative questionnaire, in which students had to demonstrate their modelling competence by understanding and applying mathematical concepts, without the requirement of performing mathematical calculations.

The experiment was then divided into four phases, as shown in Figure 3: Pre – test, Game Play, Exit Survey, and Post – test. During the Pre – test, both groups, control and experimental group, solved the argumentative questionnaire, and their outcomes were recorded. Then, for the Game Play, the experimental group played the designed game, in which the students took place of an epidemiologist facing a world pandemic. The objective of the game was to select the optimal alternatives by correctly interpreting functions' behaviors and rates of change according to the SIR model. During the Game Play phase, students entered the prompt into ChatGPT and selected what they considered the best alternatives according to their analysis of functions and rates of change. The game duration was about 15 minutes, after which they received feedback from same ChatGPT to improve their decisions, and they discussed in teams what were the improvement opportunities.

Immediately after playing the game, during phase 3 of the experiment, students answered an Exit Survey that measured their perception of Easiness of playing the game, the Interactivity achieved, how Enjoyable it was, and if their Curiosity was triggered. The instrument used was a modified version of the questionnaire used by Carrion et.al. to measure acceptance of virtual settings for learning [21]This survey allowed us to measure the acceptance of students towards the use of AI during the activity.

At the final phase, the day after the game took place, students in the experimental group solved another argumentative questionnaire, involving the same behaviors as before. Their performance was recorded to compare those results with the ones of the pre – test.

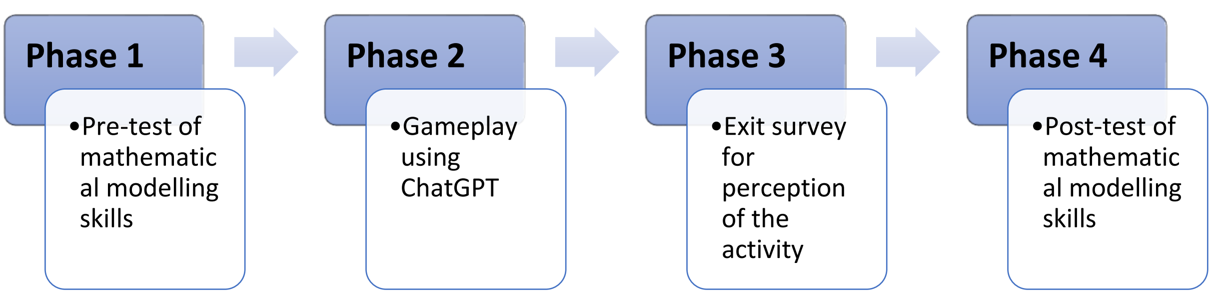

Figure 3. Implementation phases of the experiment

Discussion of Results

We analyzed our results using a 95% confidence level and hypothesis tests for difference of means. The results of the tests performed by the control group and the experimental group show two findings: First, the experimental group presents a clear shift of scores when comparing the pre-test with the post-test. Second, the control group, even it has a more disperse distribution of scores, outperformed the experimental group. Figure 4 show the a 95% confidence interval of scores in the three tests: pre-test, post-test, and control.

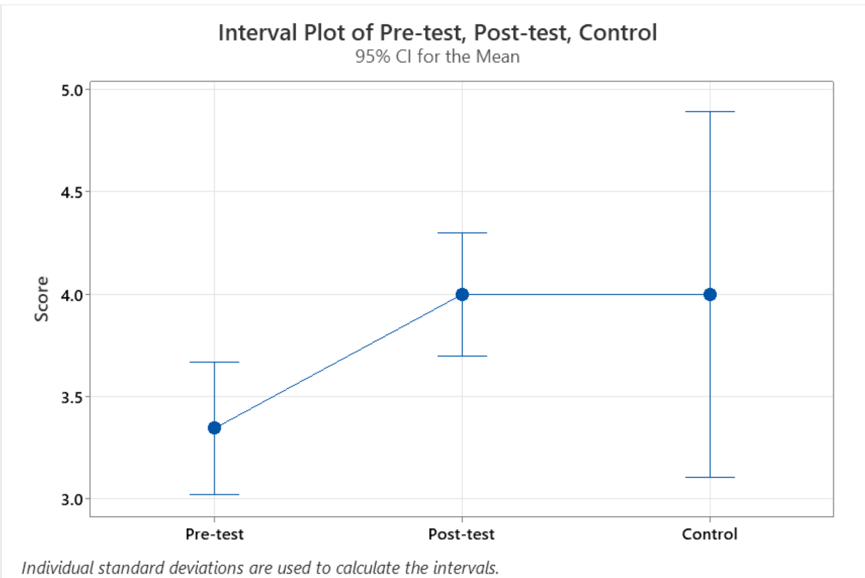

Figure 4. Confidence interval of mean scores: pre-test left, post-test middle, control right

Regarding to the first finding, a hypothesis test for paired difference of means at a 95% confidence level for the pre-test and post-test scores allowed us to see a significant improvement in students' performance in the experimental group, with an average improvement of 20%, as shown in Table 3.

Table 3. Hypothesis test for experimental group

Sample	N	Mean	Paired diff	StDev	SE	95% CI	T-Value	P-Value
Post-test	26	4.000	0.654	1.198	0.235	(0.170, 1.138)	2.78	**0.010**
Pre-test	26	3.346						

μ_difference: population mean of (post - pre)
H_0: *μ_difference = 0* H_1: *μ_difference ≠ 0*

However, regarding to the second finding, the comparison of results with the control group using a 95% hypothesis test for difference of means does not present a significant improvement in the mean score, as it can be seen in Table 4.

Table 4. Hypothesis test for experimental and control groups

Sample	N	Mean	StDev	SE Mean	Diff	95% CI	T-Value	P-Value
Post-test	26	4.000	0.748	0.15	0.000	(-0.917, 0.917)	0.00	**1.000**
Control	8	4.000	1.07	0.38				

μ_difference: population mean of (post - control)
H_0: *μ_difference = 0* H_1: *μ_difference ≠ 0*

The previous statistics clearly show that after the implementation of the proposed methodology in the experimental group, the scores increased an average of 20% and the dispersion slightly diminished in approximately 6%, as it can be seen in Figure 5. This shift to the right of the distribution allows us to infer a positive impact of the methodology in the mathematical skills of the students.

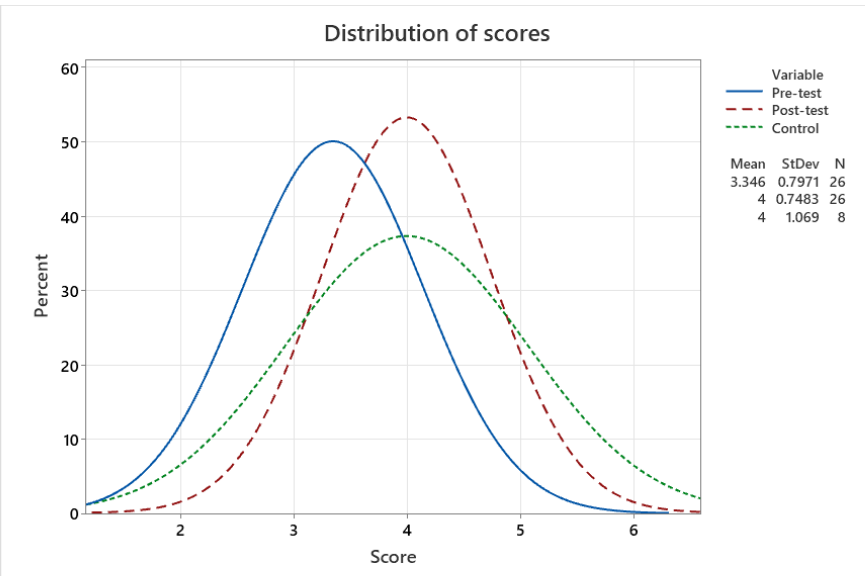

Figure 5. Distribution of scores: pre-test in blue, post-test in red, control in green

Nevertheless, since both groups were exposed to the same traditional class format, which involved theoretical explanation and exercise solving during class, the results point out a better performance of students in the control group a priori to the experiment. Further research to verify those factors that affect this outcome is necessary and will be considered in future studies.

However, to verify the reason why the experimental group did increase the mean score, we analyzed individual questions from the performed tests to compare the three mathematical skills considered in this research: Understanding the meaning of rates of change, Interpretation of graphs, and Decision making according to a mathematical model.

The questions in the tests allowed us to analyze the specific skills of the mathematical modelling competence within the experimental group. Each test contained five questions: 2 related to the skill Understanding the meaning of rates of change, 1 related to the skill Interpretation of graphs, and 2 related to the skill Decision making according to a mathematical model. An example of question, related to the first of the skills, is given below:

The temperature of a cake changes at a rate of $\frac{dT}{dt} = -2.3\,°C/min$. *Most likely, the cake...*

a) *It's in the oven*
b) *It's out of the oven*
c) *It's already at room temperature*

The comparison of the pre-test vs post-test shows that the improvement is not sustained equally among all skills. We can see at Figure 6 that there is no improvement in "Understanding the meaning of rates of change", while "Interpretation of graphs" and "Decision making according to a mathematical model" do present an improvement that can be interpreted as a better ability to use mathematical data to graphically explain a problem and to obtain a solution from it. Table 5 reinforce this inference with a 95% paired sample hypothesis test for the difference of means. It is possible to verify than only the last two skills present a significant change of mean.

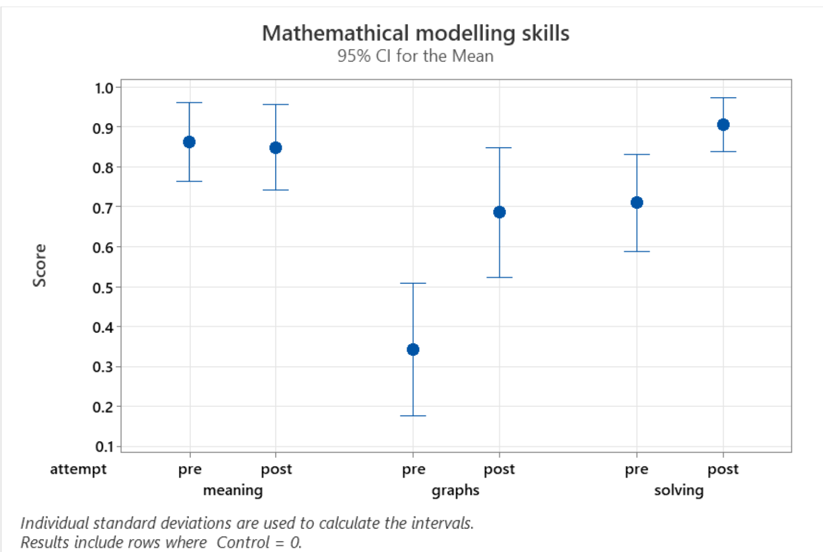

Figure 6. Improvement of mathematical modelling behaviors

Table 5. Hypothesis test for mathematical skills

Skill	Mean	StDev	SE Mean	95% Lower Bound	T-Value	P-Value
Meaning	-1.43	39.06	6.60	-12.59	-0.22	**0.585**
Graphs	34.3	63.9	10.8	16.0	3.17	**0.002***
Solving	19.52	37.81	6.39	8.72	3.05	**0.002***

μ_difference: population mean of post – pre
H_0: μ_difference = 0 H_1: μ_difference > 0
*Significant at 95%

Furthermore, the analysis of perception of students towards the AI activity indicates that the interface of the AI is easy to use and provides an acceptable level of interactivity. However, the game was not necessarily perceived as enjoyable nor generated curiosity for learning more, as seen in Figure 7. The applied questionnaire corresponds to a modified version of the instrument proposed by Carrion et.al., which measures the acceptance to technology in four dimensions: Ease of Use, Interactivity, Enjoyable experience, and Curiosity for learning. Each dimension contained different statements that the student rated from 1 to 5, where 1 corresponds to Completely Disagree and 5 to Completely Agree. It can be seen in Table 6 the questions applied to the students in the control group and their correspondent dimensions.

Table 6. Questionnaire for AI acceptance perception

Dimension	**Questions**
Easiness	Learning to use the game was easy It wasn't difficult for me to get what I needed to do in ChatGPT It was easy for me to become proficient in using GPT chat and its content
Interactivity	I found the game environment flexible to interact with I felt carried away while experiencing the activity I felt like I was part of the text game environment By experiencing the role-playing game, I felt like I am in control
Enjoyable	I enjoyed experiencing the activity Experiencing a virtual text world is quite nice The role-playing experience was fun
Curiosity	The role-playing experience is very interesting When experiencing the role-playing game, my attention was totally focused Experimenting on ChatGPT piqued my curiosity Experiencing the role-playing game was interesting

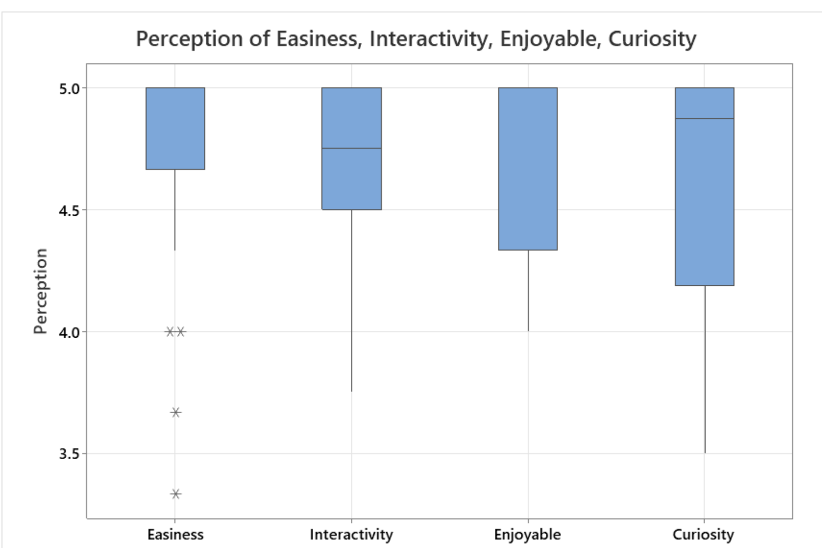

Figure 7. Perception of the students towards the activity

These results imply that further research must be performed to create activities that do increase engagement of the students. A correlation analysis of the dimensions from the perception questionnaire, shown in Table 7, points out that Enjoyment and Curiosity are highly positively correlated with Interactivity. Possibly, a prompt that allowed more interactivity for the student would improve the other dimensions as well.

Table 7. Correlation of perception survey's dimensions

	Easiness	Interactivity	Enjoyable
Interactivity	0.460*		
Enjoyable	0.288	0.601*	
Curiosity	0.289	0.823*	0.638*

**Significant correlations at a 95% confidence level*

Conclusions

The results of this study show that the use of Game – Based Learning with ChatGPT may positively improve the performance of students by contributing to the development of two mathematical skills: "Interpretation of graphs" and "Decision making according to a mathematical model". These skills, even analyzed in First Year Program students, can be beneficial for several other courses during their Engineering studies. Statistically, it was found a significant improvement when comparing the pre – test scores and the post – test scores of the experimental group, as long with a slight reduction in its variability. We can imply that the game in ChatGPT was useful to level students' competence, reducing variability of learning as expected in a homogeneous group [22].

However, the study presents that the students not necessarily perceived the AI game increased their curiosity for learning nor was sufficiently enjoyable to be played. Moreover, obtained results show a strong positive correlation between the perception of Interactivity, Enjoyable, and Curiosity. This relationship supports that introducing fun elements into instructional scheme generates engagement of students [23], and therefore contributes to the development of competences [24]. Further efforts to improve the game dynamics are necessary and will be considered in future phases of the research.

Acknowledgement

The authors would like to acknowledge the financial support of Writing Lab, Institute for the Future of Education, Tecnologico de Monterrey, Mexico, in the production of this work.

References

[1] Instituto Tecnológico y de Estudios Superiores de Monterrey, "Modelo Educativo TEC21," Monterrey, México, 2016.

[2] G. Sayeg-Sánchez, E. G. Avilés-Rabanales, and M. X. Rodríguez-Paz, "Use of Scrum in a Virtual Environment to Enhance Collaboration and Systemic Reasoning of Engineering Students," Jul. 2021.

[3] V. Barabash, M. Milz, T. Kuhn, and R. Laufer, "Development of a competence ecosystem for the future space workforce: strategies, practices and recommendations from international master programs in northern Sweden," *Acta Astronaut*, vol. 197, pp. 46–52, Aug. 2022, doi: 10.1016/j.actaastro.2022.05.017.

[4] H. van der Meij and P. Dunkel, "Effects of a review video and practice in video-based statistics training," *Comput Educ*, vol. 143, no. April 2019, p. 103665, 2020, doi: 10.1016/j.compedu.2019.103665.

[5] R. F. Herrera, M. A. Sanz, L. Montalbán-Domingo, T. García-Segura, and E. Pellicer, "Impact of Game-Based Learning on Understanding Lean Construction Principles," *Sustainability 2019, Vol. 11, Page 5294*, vol. 11, no. 19, p. 5294, Sep. 2019, doi: 10.3390/SU11195294.

[6] Y.-P. Cheng, C.-F. Lai, Y.-T. Chen, W.-S. Wang, Y.-M. Huang, and T.-T. Wu, "Enhancing student's computational thinking skills with student-generated questions strategy in a game-based learning platform," *Comput Educ*, vol. 200, p. 104794, 2023, doi: 10.1016/j.compedu.2023.104794.

[7] S. Daniau, "The Transformative Potential of Role-Playing Games—: From Play Skills to Human Skills," *Simul Gaming*, vol. 47, no. 4, pp. 423–444, Aug. 2016, doi: 10.1177/1046878116650765.

[8] P. Chan, T. Van Gerven, J. L. Dubois, and K. Bernaerts, "Study of motivation and engagement for chemical laboratory safety training with VR serious game," *Saf Sci*, vol. 167, Nov. 2023, doi: 10.1016/j.ssci.2023.106278.

[9] P. K. Parthasarathy, A. Mittal, and A. Aggarwal, "Literature Review: Learning Through Game-Based Technology Enhances Cognitive Skills," *International Journal of Professional Business Review*, vol. 8, no. 4, pp. e01415–e01415, Apr. 2023, doi: 10.26668/BUSINESSREVIEW/2023.V8I4.1415.

[10] G. Sayeg-Sánchez, M. X. Rodríguez-Paz, and D. Valencia-Marquez, "Guided Learning Sequences as an e-Learning Enhancer During COVID-19 Emergency Conditions," in *2021 ASEE Virtual Annual Conference Content Access*, Jul. 2021.

[11] M. X. Rodriguez-Paz, J. A. Gonzalez, I. Zamora-Hernandez, and G. Sayeg-Sánchez, "A hybrid online/lectures teaching model for Mechanics of Structures Courses involving new learning spaces," in *ASEE Annual Conference and Exposition*, Virtual, Online, Jun. 2020.

[12] Y.-M. Huang, M. Silitonga, and T.-T. Wu, "Applying a business simulation game in a flipped classroom to enhance engagement, learning achievement, and higher-order thinking skills," *Comput Educ*, vol. 183, no. 123, 2022, doi: 10.1016/j.compedu.2022.104494.

[13] C. Rios-Campos *et al.*, "Artificial Intelligence and Education," *South Florida Journal of Development*, vol. 4, no. 2, pp. 641–655, Apr. 2023, doi: 10.46932/SFJDV4N2-001.

[14] M. Rabeya, M. S. Mahmud, S. F. T. Johora, and A. Sattar, "A Comparison of the Application of Artificial Intelligence in the Educational System," *2022 International Conference on Emerging Trends in Computing and Engineering Applications, ETCEA 2022 - Proceedings*, pp. 1–6, Nov. 2022, doi: 10.1109/ETCEA57049.2022.10009813.

[15] C. K. Lo, "What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature," *Educ Sci (Basel)*, vol. 13, no. 4, pp. 410–410, Apr. 2023, doi: 10.3390/EDUCSCI13040410.

[16] Y. Wardat, M. A. Tashtoush, R. AlAli, and A. M. Jarrah, "ChatGPT: A revolutionary tool for teaching and learning mathematics," *Eurasia journal of mathematics, science and technology education*, vol. 19, no. 7, pp. em2286–em2286, Jul. 2023, doi: 10.29333/EJMSTE/13272.

[17] "Investigating the Effectiveness of ChatGPT in Mathematical Reasoning and Problem Solving: Evidence from the Vietnamese National High School Graduation Examination," *arXiv.org*, vol. abs/2306.06331, Jun. 2023, doi: 10.48550/ARXIV.2306.06331.

[18] L. S. Lo, "The Art and Science of Prompt Engineering: A New Literacy in the Information Age," *Internet Reference Services Quarterly*, vol. 27, no. 4, pp. 1–8, Jun. 2023, doi: 10.1080/10875301.2023.2227621.

[19] L. S. Lo, "The CLEAR path: A framework for enhancing information literacy through prompt engineering," *The Journal of Academic Librarianship*, vol. 49, no. 4, p. 102720, Jul. 2023, doi: 10.1016/J.ACALIB.2023.102720.

[20] I. A. Abelló Ugalde, R. Guinovart Díaz, and W. Morales Lezca, "El modelo SIR básico y políticas antiepidémicas de salud pública para la COVID-19 en Cuba," Rev Cubana Salud Pública 46. Accessed: Oct. 31, 2023. [Online]. Available: https://www.scielosp.org/article/rcsp/2020.v46suppl1/e2597/

[21] B. Carrion, C. A. Gonzalez-Delgado, A. Mendez-Reguera, I. E. Erana-Rojas, and M. Lopez, "Embracing virtuality: User acceptance of virtual settings for learning," *Computers & Electrical Engineering*, vol. 93, p. 107283, Jul. 2021, doi: 10.1016/j.compeleceng.2021.107283.

[22] G. Sayeg-Sanchez and M. X. Rodriguez-Paz, "Factors That Impact Mastery Learning in a Probability and Statistics Course," *Proceedings of the 2019 IEEE 11th International Conference on Engineering Education, ICEED 2019*, pp. 174–177, Nov. 2019, doi: 10.1109/ICEED47294.2019.8994920.

[23] M. J. Tews and R. A. Noe, "Does training have to be fun? A review and conceptual model of the role of fun in workplace training," *Human Resource Management Review*, vol. 29, no. 2, pp. 226–238, Jun. 2019, doi: 10.1016/J.HRMR.2017.11.003.

[24] A. Rodríguez-Muñoz, M. Antino, P. Ruiz-Zorrilla, and E. Ortega, "Positive emotions, engagement, and objective academic performance: A weekly diary study," *Learn Individ Differ*, vol. 92, no. September, 2021, doi: 10.1016/j.lindif.2021.102087.