

Article

ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics

Luis M. Sánchez-Ruiz ^{1,*}, Santiago Moll-López ¹, Adolfo Nuñez-Pérez ¹, José Antonio Morano-Fernández ¹
and Erika Vega-Fleitas ²

¹ Departamento de Matemática Aplicada, Universitat Politècnica de València, 46022 Valencia, Spain; sanmollp@mat.upv.es (S.M.-L.); adolfonunezresearch@gmail.com (A.N.-P.); jomofer@mat.upv.es (J.A.M.-F.)

² Instituto de Diseño y Fabricación, Universitat Politècnica de València, 46022 Valencia, Spain; ervefl@upv.es

* Correspondence: lmsr@mat.upv.es

Abstract: This research aims to explore the potential impact of the ChatGPT on b-learning methodologies in engineering education, specifically in mathematics. The study focuses on how the use of these artificial intelligence tools can affect the acquisition of critical thinking, problem-solving, and group work skills among students. The research also analyzes the students' perception of the reliability, usefulness, and importance of these tools in academia. The study collected data through a survey of 110 students enrolled in a Mathematics I course in BEng Aerospace Engineering where a blended methodology, including flipped teaching, escape room gamification, problem-solving, and laboratory sessions and exams with a computer algebraic system were used. The data collected were analyzed using statistical methods and tests for significance. Results indicate students have quickly adopted ChatGPT tool, exhibiting high confidence in their responses (3.4/5) and general usage in the learning process (3.61/5), alongside a positive evaluation. However, concerns arose regarding the potential impact on developing lateral competencies essential for future engineers (2.8/5). The study concludes that the use of ChatGPT in blended learning methodologies poses new challenges for education in engineering, which requires the adaptation of teaching strategies and methodologies to ensure the development of essential skills for future engineers.

Keywords: artificial intelligence; ChatGPT; GPT-3.5; GPT-4; blended learning; game-based learning



Citation: Sánchez-Ruiz, L.M.; Moll-López, S.; Nuñez-Pérez, A.; Morano-Fernández, J.A.; Vega-Fleitas, E. ChatGPT Challenges Blended Learning Methodologies in Engineering Education: A Case Study in Mathematics. *Appl. Sci.* **2023**, *13*, 6039. <https://doi.org/10.3390/app13106039>

Academic Editors: Maxim Mozgovoy and Paolo Burelli

Received: 13 April 2023

Revised: 6 May 2023

Accepted: 10 May 2023

Published: 14 May 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The emergence of generative artificial intelligence (Gen-AI) systems, or artificial intelligence (AI) in short, such as the current most popular ChatGPT tool from OpenAI [1], is becoming a significant turning point in the academic world, the consequences of which are starting to be explored [2–4], although the repercussions may be broader than anticipated. Gen-AI systems are created to produce a wide range of outputs, such as texts, images, videos, or codes, by employing a data repository that trains it. There exist some other Gen-AI systems, such as Rytr [5], Jasper [6], CopyAI [7], Writesonic [8], Kafkai [9], Copysmith [10], or Article Forge [11], but the rapid success of ChatGPT model GPT-3.5 and GPT-4 has represented a significant advancement in AI technology, which has subsequently raised concerns about their potential impact on academic integrity [2,12–19]. As stated in the literature, it is important to consider the ethical implications of these systems and plan the implementation of appropriate measures to ensure their responsible use in academic environments.

ChatGPT is a large language model in which generative pre-trained transformer models (GPT) generate content as a response to the interaction with a prompted question or command. The latest model of ChatGPT 3 was released only in November 2022, recently updated to model 3.5 in March 2023, and has spread at a dizzying speed as one of the most employed tools in the academia. GPT-4 has also been released in March 2023 improving the capabilities of GPT-3. They are designed to generate responses in dialogues/conversations to a wide variety of language tasks, and has shown, so far, more use cases than other Gen-AI systems.

GPT-3/GPT-4 has proven in a very short time to be a extremely handy tool in the academic field. The information provided by ChatGPT is being thoroughly studied and tested by OpenAI and numerous users that employ the tool to gauge its effectiveness and actual capabilities [13,14,18–23]. The ability, shown up to now, to carry out scientific essays and academic texts, as well as to solve complex problems, has called into question the efficiency of different learning methodologies and the appropriate use of these tools [3,13,17,23]. Since ChatGPT facilitates or directly solves the tasks posed by the teachers, it may enable students to bypass the learning process and acquire answers without developing the necessary knowledge, skills, or competencies [3,13,23].

For more than a decade, following the Joint Declaration of the European Ministers of Education convened in Bologna on 19 June 1999 [24], universities began a process where the teacher-centered model gave way to a new model, where the weight of learning was based more on students' needs and paces. This evolution was accompanied by a more widespread use of digital tools, which were increasingly being implemented in the classroom and in the learning process. These two facts favored the rise of new educational methodologies, such as blended learning (BL or b-learning), which combines traditional face-to-face instruction with various digital technologies and resources. These strategies allow for greater flexibility in how, when, and where students learn, promoting the creation of a customized learning experience.

These strategies, which were progressively implemented in the academic world, helped to soften the impact of COVID-19 on education. The lockdown and the shift from a mixed face-to-face and online teaching (in the best cases) to a fully online environment caused a sudden change and an immediate need for technological adaptation. Literature can be found discussing different approaches and evaluating the consequences of this adaptation, see, for example, [25–29]. After the pandemic, many digital elements remained, while others slowly faded away. These remaining tools, such as online exams, quizzes, knowledge reinforcement exercises, and games are mainly implemented, as in the case we are addressing, in blended learning methodologies. In this article, we will focus on examining the possible consequences of introducing AI into these types of methodologies, as they seem to have more vulnerabilities when faced with such tools.

Blended learning can be defined as a student-centered approach that combines the benefits of online learning (flexibility, abundant resources, and timely updates) with the interactivity of traditional teaching. Researchers have assessed the feasibility and effectiveness of these models through multiple dimensions, such as knowledge acquisition, competencies performance, technology availability, and satisfaction. Although the concept of BL is not new, its use and implementation have been expanding in the academic world [30–32]. This growth has been driven and reinforced by the previously mentioned technological advancements and the extensive range of resources that have been introduced.

BL integrates the best aspects of traditional face-to-face instruction with online learning components to create an optimal, flexible, and engaging learning experience for students [29,33,34]. This educational methodology has demonstrated several positive outcomes in education, such as improved learning outcomes, increased student commitment and satisfaction, enhanced self-regulated learning and time management, increased access and adaptability, and cost-effectiveness [35–43].

In a BL environment, students can benefit from various learning modalities, from face-to-face instruction (involving in-person interactions between teachers and students, direct communication and immediate feedback [44]), online learning (involving self-paced learning through online resources), and collaborative learning (encouraging collaboration among students and knowledge sharing). However, some studies have reported potential challenges or negative outcomes, such as technology barriers that can hinder the effectiveness of BL for some students [45], an increased workload for educators [46], social isolation [47] or difficulty in adapting to the learning format [48]. These challenges can be addressed through careful planning, providing adequate support to both students and educators, and continuously evaluating and refining the BL approach.

In blended learning, many other methodologies can be included, such as flip-teaching (FT) or the use of game-based learning (GBL). However, since these latter methodologies have their own distinctive elements, we treat them as separate methodologies. These two learning methodologies that have shown considerable success in academia, and may also be affected by the emergence of AI technologies.

FT methodology aims to foster active participation in the learning process by incorporating out-of-classroom activities. These activities are designed to help students learn, practice, and master the required concepts. Recent studies have shown the effectiveness of FT in promoting student engagement, improving student performance and satisfaction, and enhancing critical thinking skills [49–52].

Another of the methodologies in the BL setting used in this research has been GBL. This methodology involves introducing games or game elements in the classroom, in order to promote motivation, active participation, and the creation of a positive learning environment [53–61]. Among the games that can be introduced in the classroom, Escape Rooms (ER) have recently gained significant popularity in the academic world, due to their adaptability to different environments, promotion of collaborative work, and varied nature of the challenges to be solved. When this type of game is applied with the aim of promoting learning and competency development in education, they are called Educational Escape Rooms (EERs) [62–64]. EERs are usually developed in the classroom or in a controlled environment with physical participation of students in groups. There is also a digital version of these games (digital Educational Escape Rooms, dEERs), which can be collaborative or not, and which students have the option to do outside of the classroom. This modality was motivated by the simplification of resources and the COVID-19 pandemic [29,65,66]. The application of this type of game seems to be related to promoting students' learning process and enhancing the development of transversal competencies, such as team-work, lateral and critical thinking, communication, and working under pressure, among others [62,64–69]. Escape rooms are based on implementing a theme and a narrative that serves as the guiding thread of the activity. The tremendous thematic variety allows these dEER to be applied in many contexts.

AI, specifically ChatGPT, has demonstrated a high level of proficiency in composing texts and essays, translating between various languages, and generating original ideas. In Science, Technology, Engineering and Mathematics (STEM) area, this tool, in addition to aiding the aforementioned activities, faces the added challenge of performing calculations and solving scientific problems, such as engineering issues or mathematical challenges. It is in this area where more difficulties arise in obtaining reliable answers. The mathematics subject, which we focus on in this study, requires an advanced understanding of certain concepts and, most importantly, the prior development of specific competencies for success. The use of scientific language, which aids in understanding and solving problems that arise in STEM once mastered, is also a cornerstone in the smooth development of the learning process. Many students exhibit weaknesses in some competencies, especially in the early years of their studies, so an appropriate course design can help address these shortcomings. The emergence of tools such as ChatGPT, which could potentially solve these problems, might weaken the learning process by hindering the deep assimilation of techniques and results.

All three methodologies (BL, FT, GBL) aim to increase student engagement and improve learning outcomes by making learning more interactive, flexible, and enjoyable for students. While each method has its unique approach, they can be combined or used interchangeably, depending on the needs of the students and the goals of the educators. Nevertheless, they can be threatened or reinforced by the emergence of ChatGPT. When faced with a challenge or problem, students typically seek help from teachers, consult online resources, such as web pages, texts, videos, or tutorials, and fill in gaps in their knowledge. However, the emergence of ChatGPT presents a new challenge to this process, potentially diminishing the efficiency of existing teaching methodologies and the designed activities [3,14].

This article seeks to evaluate ChatGPT models GPT-3.5 and GPT-4 problem-solving capabilities in a mathematical context, specifically within the STEM field. The study focuses on Mathematics I, a course offered at the Higher Technical School of Design Engineering, Technical University of Valencia (UPV), Spain. The course employs a BL methodology, where laboratory work and weekly tasks are conducted using a FT approach, while other methodologies, such as GBL techniques, are utilized in the classroom setting [44]. By testing its abilities in a real-world academic setting, and studying students' potential use and opinions, this study can provide valuable insights into the possible consequences that ChatGPT and other generative AI tools can have on the BL methodology applied in this case study and the implications in general STEM education.

2. Materials and Methods

2.1. Questionnaire

A 15-question ad hoc questionnaire (see Table 1) has been designed for data collection. The validity of the questionnaire designed to study the students' uses and opinions toward ChatGPT has been checked by subject area experts. Answers used 5 Likert scales [70].

The questionnaire used in this study demonstrated good internal consistency, as evidenced by a Cronbach's alpha of 0.78. This indicates that the instrument is reliable for assessing the students' uses and opinions toward ChatGPT and the questionnaire ensures that the collected data are consistent and suitable for further analysis. The questionnaire was also reviewed and approved by a group of 4 subject area experts to ensure content validity.

The questionnaire was disseminated via Typeform [71]. Ensuring complete anonymity, the surveys allowed for voluntary participation, and students were free to discontinue completion at any point.

The questionnaire used in this study has been organized into the following streamlined dimensions:

1. **Demographic data:** This dimension collects demographic information about the participants' gender, which can be helpful in understanding potential differences in the experiences and perspectives of male and female students.
2. **Usage patterns of ChatGPT:** This dimension explores both the frequency and specific situations in which students use ChatGPT, encompassing its application in academic coursework, as well as out-of-classroom and digital activities in a BL setting.
3. **Students' opinions on ChatGPT:** This dimension investigates students' perceptions of the reliability and trustworthiness of ChatGPT, as well as its role and impact in the educational learning process.
4. **Impact on competencies:** This dimension assesses the perceived impact of ChatGPT on students' competencies development.

Following the survey, a group of volunteer students participated in interviews to address questions concerning the use of ChatGPT. A structured interview guide was created, drawing upon the questionnaire items. Each interview consisted of 10 questions and took less than 5 min to complete. Only 10 students volunteered to be interviewed. Comprehensive notes were documented during the interviews, capturing the students' responses and observations.

Table 1. Questions in the questionnaire.

	Question	Answers
Q1	Gender	Male/Female/Empty
Q2	Do you know what ChatGPT is?	Yes/No
Q3	Do you use it for academic purposes?	Yes/No
Q4	When did you start using it?	Nov–Mar
Q5	How often do you use the ChatGPT tool for academic purposes?	Likert scale 1–5
Q6	How often do you use the ChatGPT tool in Mathematics I?	Likert scale 1–5
Q7	How reliable do you think the answers of ChatGPT are with respect to the theoretical mathematical background?	Likert scale 1–5
Q8	How reliable do you think the answers of ChatGPT are with respect to numerical calculations?	Likert scale 1–5
Q9	Do you think that the use of ChatGPT has helped you to learn/reinforce some mathematical concepts used in Mathematics I?	Likert scale 1–5
Q10	Do you think that the use of ChatGPT has helped you in solving problems/exercises in Mathematics I?	Likert scale 1–5
Q11	Have you used ChatGPT to help you complete scheduled tasks outside the classroom?	Likert scale 1–5
Q12	Do you think ChatGPT could be an important tool in academia?	Likert scale 1–5
Q13	Do you think using ChatGPT could affect the acquisition of problem-solving competency?	Likert scale 1–5
Q14	Do you think using ChatGPT could affect the acquisition of critical-thinking competency?	Likert scale 1–5
Q15	Do you think using ChatGPT could affect the acquisition of group work competency?	Likert scale 1–5

2.2. Sample

The sample consists of fresh students from a Mathematics I subject in Aerospace Engineering within a technological university (Universitat Politècnica de Valencia, Valencia, Spain) in which the authors are involved as instructors. This makes it a convenience sample where students are accustomed to blended and digital methodologies. Additionally, 20% of the students are female, which matches the general proportion at the technological university where these studies are developed, primarily focusing on STEM subjects.

The subject has 128 freshmen engineering students enrolled. Out of these, 110 students completed the survey during the 2022/2023 academic year. The students' ages ranged from 18 to 19 years old. Data collection took place at the beginning of March 2023.

As limitations of the sample, it is worth noting a limited generalizability to first-year engineering courses with a fairly specific student profile. This study aims to be a first step in the more general examination of opinions and uses of AI by students in the STEM area.

2.3. Data Analysis

The data analysis and treatment were performed using SPSS software and Excel software was utilized for generating graphs. Normality testing of the data was conducted using Shapiro–Wilk and Kolmogorov–Smirnov tests [72]. Results from both tests indicated non-normal distribution of questionnaire scores, therefore the Mann–Whitney U Test non-parametric statistical method was applied when needed.

For the study on students' opinions and usage, only ChatGPT model 3.5 has been taken into account as it is a free tool. GPT-4 has only been evaluated by the authors as, for now, it is a subscription-based tool.

During the data analysis phase, there was no need for a data cleaning process, as the collected responses were complete, consistent, and free of any apparent discrepancies or outliers.

2.4. BL Setting

The BL approach is structured around three stages, autonomous learning, online knowledge assessment, and in-class reinforcement (see [29]). Autonomous Learning involves students accessing resources and learning at their own pace outside the classroom through the PoliformaT platform (Sakai). Online knowledge assessment, the second stage, aims to assess students' learning through online tests, providing both students and teachers with immediate feedback. Finally, in-class reinforcement focuses on interactive, collaborative, and activity-based strategies to reinforce knowledge and competencies acquired during the first two stages. This approach maximizes student–teacher interaction and helps students better engage with the material, leading to more precise and targeted questions and discussions.

A novel assessment method called Dynamical Continuous Discrete Assessment (DCDA) (see [44]) that aims to evaluate individual student progress and enhance formative capabilities by helping and encouraging students to reach and improve their expected competencies throughout the course is also applied. The DCDA system builds upon the existing Continuous Assessment (CA) paradigm and combines it with a discrete dynamical approach to consider the interconnectedness of course topics. This method acknowledges that each assessment and output is an input into the learning process and requires subsequent assessments to confirm or reassess the level of competencies achieved. By integrating DCDA into the three-stage blended learning methodology based on Flipped Teaching, the authors aim to create a more comprehensive and effective learning experience for students.

2.5. Digital Educational Escape Rooms and Data Collection

Five short-duration dEERs were designed for different topics in the syllabus of Mathematics I, with 3 implemented for algebra and 2 integral calculus. Each dEER aimed to reinforce knowledge and was implemented and designed using RPG Maker MZ [73] software. They were based on a specific narrative (science fiction and fantasy) with different challenges to be overcome by solving mathematical puzzles and problems in a limited time. They were designed to be played collaboratively in groups of 4–5 students, but were also available for individual play.

The dEERs puzzles had a linear structure consisting in escaping different rooms or levels, with game mechanics found in digital games. Solving the puzzles required direct answers based on a specific mathematical knowledge and competencies, or they involved finding hidden symbols or solving numerical problems with a high degree of accuracy. The levels were designed to be challenging, with most rooms having 3–4 problems and 2–3 tests. Failed attempts were sometime penalized on the avatars' personal characteristics (life points or strength points). The tests could be simple-choice or multiple-choice questions, and they could include graphical answers (see Figure 1).



Figure 1. Images of different games and quizzes designed with RPG Maker. Maps are modifications of RPG Maker MZ library.

The use of these digital games provided the authors with data on student performance and knowledge. The game collects data, such as response time, number of failed attempts, provided answers, hours of access, and opinions. However, it does not collect personal information of the students, so it is not possible to identify a pattern of responses with a specific student, only with an avatar. This information is used as a tool for a feedforward

strategy that allows the student to address mathematical and competency deficiencies before the evaluation [44].

3. Results

3.1. Mathematical Tests with ChatGPT

Before the lab session, students engage in self-directed learning using the FT methodology. This involves accessing exercises and explanatory texts with examples similar to those they will face in the lab session. At the start of the class, the teacher reviews the content, and students have the opportunity to ask questions. The lab sessions are based on the use of Wolfram Mathematica software [74], version 12 or 13, a highly reliable and powerful mathematical calculation tool capable of solving a wide range of physical and mathematical problems.

However, using Mathematica requires prior knowledge of the correct syntax and the appropriate commands to use for each calculation. To help with this, Mathematica provides an assistant that suggests and corrects possible errors in syntax. Once the command is entered correctly, Mathematica returns a solution with an impressive level of reliability and precision.

One possible disadvantage of using Mathematica is that it needs some mathematical background to identify the problem to be solved, using the proper command or function, and interpreting correctly the outputs. Mathematica software provides a vast digital library with thousands of examples that students can access through the Wolfram Documentation Assistant to help them learn the fundamental mathematical knowledge and the syntax and command structure necessary for solving problems. Additionally, the UPV offers free access computer rooms in all schools and faculties; Wi-Fi technology in its three campuses, both in buildings and in gardens and outdoor areas, and its Mathematica license enables its use available to current faculty, staff and students for teaching, learning and academic research, and even to install it in their laptops with a provisional 1-year renewable licence. For those who do not have access to Mathematica, Wolfram Alpha [75] is a free tool that can be used to solve a wide range of mathematical problems, using a more general syntax. Wolfram Alpha can be accessed via the internet, and calculations are performed on the Wolfram server rather than on personal computer hardware.

In the laboratory sessions of Mathematics I, Wolfram Mathematica is a fundamental tool for solving exercises. Students need to review the theoretical knowledge and learn the specific syntax of the software to solve the problems successfully. During the session, after the teacher's explanation and doubt clarification, students take a test that includes exercises similar to those reviewed at home. The tests are carried out on a weekly basis, and in each session, different mathematical topics related to the theoretical concepts covered in the theory sessions are addressed. This paper shows the results of 18 tests, from Test01 to Test18. The evaluated tests cover the topics of complex numbers, hyperbolic functions, root finding, calculus of integrals, applications of integral calculus, numerical integration, improper integrals, systems of equations, matrices, determinants, curve fitting, vector spaces, Euclidean spaces, linear applications, and matrix diagonalization. The tests are designed to be carried out within a limited time and in a controlled environment. Despite having access to the Internet during the session, students are expected to behave honestly during the test. The aim of this weekly learning process is to reinforce the critical thinking skills of the students and to gain a deep understanding of the mathematical concepts required in a Bachelor's degree in engineering. The advent of ChatGPT has raised questions about its impact on this students' learning process. If misused, it could lead to an impoverishment in the acquisition of competencies, while, if used correctly, it could reinforce the mathematical knowledge. The authors have attempted to solve these tests with ChatGPT in order to evaluate GPT-3.5 (Legacy) and GPT-4 capabilities in solving the problems presented in the laboratory sessions.

Unlike Mathematica, ChatGPT interface offers a much more flexible syntax for requesting calculations. Students can simply copy and paste the problem into the system,

and ChatGPT will evaluate the problem and determine how to solve it. This approach has shown remarkable reliability, with a success rate of 96% for model GPT-3.5 and 98% for model GPT-4 in interpreting the meaning of a collection of 100 mathematical exercises covering various problems in Differential Calculus, Algebra, Integral Calculus, and Series, and offering an appropriate theoretical answer. We assigned a score of 1 for correct interpretations and 0 for incorrect ones to measure the reliability. It is important to note that the way in which the question is written can affect the system's understanding of the problem, so this indicator should only be taken as a rough estimate of the reliability of the answer. In the few cases where ChatGPT failed to interpret the problem correctly, the appropriate answer was obtained after no more than two interactions with the system. In addition, ChatGPT provides a detailed response with the necessary steps to solve the problem, which is a significant advantage during its use.

Next, we evaluate the accuracy of ChatGPT's numerical mathematical solutions (NMS). For example, when asked to diagonalize a matrix, part of the problem involves correctly identifying the task at hand, while the other part involves correctly solving the problem numerically and providing the appropriate matrices. GPT-3.5 has shown lower accuracy in this second part of the calculation (see Table 2).

Initially, only 36% of the problems were solved correctly. However, when an error was detected and corrected through interaction with the AI, the success rate slightly increased. For instance, a dot product of two vectors repeatedly yielded errors, but after a third or fourth interaction, the success rate increased to 44%. It is important to note that ChatGPT offers the option to re-evaluate the answer without indicating any reason, using an equivalent technique if possible. This showcases the AI's impressive versatility, but it also means that the reassessed answer may not always be accurate, despite the initial solution being correct. One question that arises from these findings is whether ChatGPT is capable of passing tests without the intervention of the user.

The results suggest that while ChatGPT still struggles with passing exams primarily based on mathematical calculations, it does demonstrate remarkable proficiency in theoretically orienting the posed problem. Therefore, it can be a valuable resource for students during their learning activities. Furthermore, it is worth noting that even when ChatGPT does not provide a correct theoretical answer, students with prior knowledge of the subject can leverage their critical thinking skills to rephrase or break down the question into smaller parts so that ChatGPT can provide satisfactory answers. This indicates that ChatGPT can function as a complementary tool to traditional learning methods rather than a substitute.

As can be seen in Table 2, the reliability of AI with respect to Theoretical Mathematical Solution (TMS) is extremely accurate; in 90% of the tests, model GPT-3.5 has provided the correct solution to the problem, although it has failed in the calculations performed. GPT-4 improves the theoretical results up to 95%, although it also fails in the calculations. However, during the process, an improvement in the final results can be appreciated, since GPT-4 increases the score obtained in 70% of the occasions. Although it cannot yet be claimed that they are capable of passing a purely numerical exam, it can be said without any doubt that they have been able to understand and provide the necessary steps for their resolution. In fact, the authors have conducted the experiment of solving the problems following the steps indicated by GPT-3.5 and GPT-4 but performing the calculations with Mathematica, and the results have been extremely good; all the tests and exam obtained more than a 8.5 (Mean = 9.5, Median = 9, SD = 1.5).

Table 2. Results from GPT-3.5/GPT-4 models in terms of the Theoretical Mathematical Solution (TMS) and Numerical Mathematical Solution (NMS).

Test	TMS GPT-3.5	NMS GPT-3.5	Score GPT-3.5	TMS GPT-4	NMS GPT-4	Score GPT-4
Test00	Correct	Correct	4.00	Correct	Correct	5.00
Test01	Correct	Incorrect	3.00	Correct	Incorrect	3.00
Test02	Correct	Incorrect	3.00	Correct	Incorrect	3.00
Test03	Correct	Incorrect	2.25	Correct	Incorrect	4.25
Test04	Correct	Incorrect	3.25	Correct	Incorrect	4.00
Test05	Correct	Incorrect	1.25	Correct	Incorrect	2.50
Test06	Correct	Incorrect	2.00	Correct	Incorrect	2.50
Test07	Correct	Incorrect	0.00	Correct	Incorrect	2.00
Test08	Correct	Incorrect	3.00	Correct	Incorrect	3.00
Test09	Correct	Incorrect	4.00	Correct	Incorrect	4.25
Test10	Correct	Incorrect	2.00	Correct	Incorrect	2.25
Test11	Correct	Incorrect	1.50	Correct	Incorrect	2.75
Test12	Correct	Incorrect	0.00	Correct	Incorrect	3.00
Test13	Correct	Incorrect	1.25	Correct	Incorrect	1.25
Test14	Correct	Incorrect	4.50	Correct	Incorrect	4.50
Test15	Correct	Incorrect	4.25	Correct	Incorrect	5.35
Test16	Correct	Incorrect	5.40	Correct	Incorrect	6.25
Test17	Correct	Incorrect	5.50	Correct	Incorrect	6.25
Test18	Correct	Incorrect	4.25	Correct	Incorrect	6.35
TestC1	Incorrect	Incorrect	3.50	Incorrect	Incorrect	4.25
TestA1	Incorrect	Incorrect	4.00	Correct	Incorrect	4.50

3.2. Digital Escape Rooms and ChatGPT

This subsection examines ChatGPT's problem-solving abilities on the dEERs designed for the course. ChatGPT was applied for the resolution of 5 dEERs that covered 3 for algebra (dEER1, dEER2, and dEER3) and 2 for integral calculus (dEER4 and dEER5). The concepts were related to those seen in the corresponding parts of the theory. In each of the dEERs, the tests were of two types, numerical problem-solving and multiple-choice questions with different response options. ChatGPT's performance in both types of questions is different due to the nature of the questions.

The responses to the numerical problem-based questions were similar to those obtained in the laboratory session tests since they are based on numerical results with considerable precision. However, due to the fact that Mathematica was not required during the game, and, therefore, the questions were designed to not require a very powerful calculation engine, the number of correct responses increased. ChatGPT (both GPT-3.5 and GPT-4) performed well in the multiple-choice questions, with better results (see Table 3). This could have important implications for the use of these methodologies and the reduction in competencies that need to be reinforced.

In this case, we did not evaluate response by response. The problem was presented directly to ChatGPT. When the response was incorrect, the problem was presented again, as the game allows for multiple attempts (although there is a limit due to penalties around of seven tries). For this reason, the performance is evaluated only based on whether ChatGPT was able to complete the dEER (success) or not (failure), implying that either it succeeded on some attempt or failed on all (Table 3).

Table 3. Results from GPT-3.5/GPT-4 models when applied to dEERs.

dEER	Score GPT-3.5	Score GPT-4
dEER1 (Algebra)	Failure	Success
dEER2 (Algebra)	Failure	Failure
dEER3 (Algebra)	Success	Success
dEER4 (Integral Calculus)	Success	Success
dEER5 (Integral Calculus)	Failure	Success

3.3. Students' Opinion and Use

This subsection analyzes the data collected from surveys regarding the usage and opinions of students in the Mathematics I course. The survey was conducted among 110 out of 128 enrolled students, which represents a high participation rate of around 86%. Among the respondents, 74.5% identified themselves as male, while 25.5% identified as female.

This study begins by analyzing how quickly this tool has spread among students (question Q4 of the questionnaire, see Table 1). The results show that ChatGPT has been widely adopted since its release in November 2022. All surveyed students reported being aware of the ChatGPT tool, and approximately 70% started using it for academic purposes in January, as shown in Table 4. This highlights the significant impact of ChatGPT in the academic community and its widespread adoption among students.

Table 4. ChatGPT usage starting time.

Month	Frequency	Percent	Cumulative Percent
November 2022	0	0%	0%
December 2022	17	15.5%	15.5%
January 2023	76	69.1%	84.6%
February 2023	15	13.6%	98.2%
March 2023	2	1.8%	100.0%

Next, the study examines how frequently students use ChatGPT in the general academic context (Q5) and in the context of the mathematics subject in particular (Q6). The aim is to assess whether students use this tool more or less in the subject under study compared to other academic activities. Regarding the use of the ChatGPT tool for academic purposes, the answers varied from “(1) I do not use it at all”, “(2) I use it very rarely”, “(3) Occasionally”, “(4) Quite often”, and “(5) I use it a lot”. The results showed that students used the tool quite frequently (Mean = 3.06, Median = 3, SD = 1.30). When considering the gender, results were similar for male/female respondents, as shown in Table 5. Although women tended to use the tool more often than men, the difference was not statistically significant (p -value = 0.09).

When evaluating the use in the Mathematics I subject, (see Table 5, fifth and sixth columns), it can be observed that the average decreases for both men and women compared to general use in the academic context. However, when comparing the means of general use with the use in Mathematics I, there is no significant difference between the means (p -value = 0.1, paired sample t -test).

After studying the frequency of use, and seeing that the use of this tool is quite widespread and therefore seems to constitute another tool in the students' learning process, we wanted to evaluate how much credibility they give to ChatGPT in two separate areas. On one hand, the theoretical mathematical response, in which it explains the concepts involved, and on the other hand, the computational aspect, in which numerical answers to the problems posed are provided. The responses to the question “How reliable do you think the answers of ChatGPT are with respect to the theoretical mathematical background?” were recorded on a 5-point Likert scale ranging from 1 (not at all reliable) to 5 (very reliable), and were collected both overall and stratified by gender.

Table 5. Statistics of use frequency considering gender as a factor.

Gender		Statistic General	Std. Error	Statistic Mathematics I	Std. Error
Female	Mean	3.43	0.202	2.89	0.23
	Median	4.00		3.00	
	Variance	1.14		1.51	
	Std. Deviation	1.07		1.23	
	Skewness	−0.39	0.44	0.09	0.44
	Kurtosis	−0.48	0.86	−0.84	0.86
Male	Mean	2.94	0.15	2.72	0.12
	Median	3.00		3.00	
	Variance	1.84		1.27	
	Std. Deviation	1.35		1.13	
	Skewness	−0.01	0.27	0.42	0.27
	Kurtosis	−1.12	0.53	−0.49	0.53

Overall, the confidence in the mathematical background of the ChatGPT responses was found to be very high (Mean = 4.21, Median = 4, SD = 0.73), with a fairly low standard deviation (see Table 6). As far as the authors have been able to verify, the ChatGPT responses regarding the problems at hand have been very accurate, without considering the calculations, and are capable of providing a fairly reliable step-by-step guide.

However, when comparing the confidence means between men and women, a slight but significant difference was observed (p -value = 0.024). Specifically, women expressed slightly lower confidence levels than men (see third and fourth columns of Table 6).

Table 6. Statistics of confidence regarding mathematical theoretical content with gender as a factor.

Gender		Statistic Theory	Std. Error	Statistic Calculations	Std. Error
Female	Mean	3.93	0.15	2.29	0.22
	Median	4.00		2.00	
	Variance	0.59		1.32	
	Std. Deviation	0.77		1.15	
	Skewness	−0.41	0.44	0.49	0.44
	Kurtosis	0.15	0.86	−0.54	0.86
Male	Mean	4.31	0.08	2.38	0.11
	Median	4.00		2.00	
	Variance	0.49		1.00	
	Std. Deviation	0.70		1.00	
	Skewness	−0.50	0.27	0.38	0.27
	Kurtosis	−0.83	0.53	−0.58	0.53

From Table 6, we infer that the confidence in the computational aspect of ChatGPT's answers is not as high as in the theoretical one. Indeed, a significant difference was found between the means obtained in confidence in the theoretical and calculistic aspects of ChatGPT (p -value = 0.001, Independent Sample T Test). However, when studying the difference between the means of the responses of men and women in terms of the reliability of the calculations, no significant difference was found (p -value = 0.617).

After examining general use and reliability, we now analyze the usefulness of ChatGPT in fostering the learning of mathematical concepts. The survey question Q9 asked: "Do you think that the use of ChatGPT has helped you to learn/reinforce some mathematical concepts used in the subject of Mathematics I?" Responses ranged from 1 (no, it has not helped me) to 5 (yes, a lot of times).

Responses (see Table 7) show a positive appreciation of the usefulness of ChatGPT in learning or reinforcing mathematical concepts (Mean = 3.50, Median = 4.00, SD = 1.03). When considering gender, the mean for men (Mean = 3.46) and women (Mean = 3.61) did not differ significantly (p -value = 0.506).

Table 7 also shows the results related to the question Q10: “Do you think that the use of ChatGPT has helped you in solving problems/exercises in the subject of Mathematics I?” Responses ranged from 1 (no, it has not helped me) to 5 (many times).

Responses also showed a positive appreciation of the usefulness of ChatGPT in solving mathematics problems/exercises (Mean = 3.37, Median = 3.00, SD = 1.19). When considering gender, the mean for men (Mean = 3.35) and women (Mean = 3.43) did not differ significantly (p -value = 0.813).

As observed in Table 7, the means did not differ much between the responses, indicating that students found ChatGPT responses quite useful in the learning process and in solving problems. This seems to indicate that, despite the short time it has been in use, students have already integrated it into their digital learning environment.

Table 7. Statistics of utility regarding mathematical learning process and problem-solving with gender as a factor.

Gender		Statistic Learning	Std. Error	Statistic Problems	Std. Error
Female	Mean	3.61	0.21	3.43	0.21
	Median	4.00		3.00	
	Variance	1.21		1.22	
	Std. Deviation	1.10		1.10	
	Skewness	−0.38	0.44	0.28	0.43
	Kurtosis	−0.40	0.86	−1.23	0.86
Male	Mean	3.46	0.11	3.35	0.14
	Median	3.5		3.00	
	Variance	1.02		1.51	
	Std. Deviation	1.00		1.23	
	Skewness	−0.23	0.27	−0.02	0.27
	Kurtosis	−0.48	0.53	−0.99	0.53

Once the usefulness of ChatGPT in students’ learning process has been established, it is logical to ask to what extent they use it, not only to improve this learning process, but also to address doubts in tasks and exercises that are part of a BL methodology structure. This is the most delicate part, as the activities, especially those planned to reinforce students’ critical thinking, can be affected by a tool that provides answers and reasoning without the student properly assimilating them in a not controlled environment. To assess students’ use of ChatGPT for completing tasks and assignments, they were asked if they had used ChatGPT to help them complete scheduled tasks outside the classroom (Q11). Responses ranged from 1 (no, never) to 5 (yes, many times) and are summarize in Table 8. The mean values obtained from the responses were lower than those in other categories (Mean = 2.33, Median = 2, SD = 0.97). However, caution must be exercised when interpreting these responses as the neutral tone of the question may have caused students to infer a search for information about a possible misuse of ChatGPT. Descriptors based on gender can be found in Table 8.

Table 8. Statistics of utility regarding the use of ChatGPT in out-of-the-class activities with gender as a factor.

Gender		Statistic Use B-Learning	Std. Error
Female	Mean	2.68	0.21
	Median	2.50	
	Variance	1.19	
	Std. Deviation	1.10	
	Skewness	0.52	0.44
	Kurtosis	−0.21	0.86
Male	Mean	2.21	0.10
	Median	2.00	
	Variance	0.81	
	Std. Deviation	0.89	
	Skewness	0.51	0.27
	Kurtosis	0.13	0.57

Students were surveyed about the importance of AI in academia (Q12), with responses ranging from 1 (not at all important) to 5 (very important). Results indicated that students generally considered these tools to be important in the academic world (Mean = 3.78, Median = 4, SD = 0.95). Table 9 presents the results stratified by gender. A significant difference was found between the responses of men and women (p -value = 0.002), with men giving greater importance to these tools than women. Table 9 summarizes the students' opinion results on how important the new tool is in academia, underscoring the increasing rapidity in which it has been integrated and its potential in this area.

Table 9. Statistics of importance of ChatGPT in academia with gender as a factor.

Gender		Statistic Importance	Std. Error
Female	Mean	3.25	0.21
	Median	3.50	
	Variance	1.23	
	Std. Deviation	1.11	
	Skewness	−0.19	0.44
	Kurtosis	−1.02	0.86
Male	Mean	3.96	0.09
	Median	4.00	
	Variance	0.68	
	Std. Deviation	0.82	
	Skewness	−1.02	0.27
	Kurtosis	1.75	0.57

One concern regarding the use of ChatGPT is whether it will hinder students' acquisition of essential skills in the development of their coursework. In this section, we evaluate students' opinions on three competencies critical to their academic development, critical thinking (CT), problem-solving (PS), and group work (GW). Responses to these competencies included the following options: (1) no, it will not affect at all, (2) yes, it will affect very little, (3) yes, it will affect somewhat, (4) yes, it will quite affect, and (5) it will affect a lot. The answers varied depending on the competency being evaluated:

- Critical Thinking: Mean = 2.38, Median = 2, SD = 1.10.
- Problem-solving: Mean = 2.39, Median = 2, SD = 1.28.
- Group work: Mean = 2.97, Median = 3, SD = 0.83.

The responses indicate that students perceive ChatGPT as having a small to moderate effect on the acquisition of the aforementioned competencies. Group work appears to be the most affected competency, according to the opinions of the students.

Table 10 shows the values of these opinions based on gender, providing a sense of how students believe that using AI affects the acquisition of competencies. Significant differences were found in students' perceptions of how ChatGPT impacts their problem-solving skills (p -value = 0.008) and critical thinking (p -value = 0.017). However, there is greater consensus on how it will affect group work (p -value = 0.687).

Table 10. Statistics of effects on competencies with gender as a factor.

Gender		Statistic PS	Std. Error	Statistic CT	Std. Error	Statistic GW	Std. Error
Female	Mean	2.89	0.23	2.86	0.24	2.89	0.19
	Median	3.00		3.00		3.00	
	Variance	1.43		1.61		0.99	
	Std. Deviation	1.26		1.27		0.99	
	Skewness	0.88	0.27	0.29	0.44	0.67	0.26
	Kurtosis	−0.33	0.53	−0.81	0.86	0.52	0.52
Male	Mean	2.21	0.14	2.22	0.11	3.00	0.09
	Median	2.00		2.00		3.00	
	Variance	1.60		0.96		0.59	
	Std. Deviation	1.26		0.98		0.77	
	Skewness	0.88	0.27	0.67	0.27	0.67	0.26
	Kurtosis	−0.33	0.53	0.20	0.53	0.52	0.52

3.4. Initial Results on Performance

In this section, the students' results are compared with those from previous years with the aim of examining significant differences in competencies acquisition.

Table 11 displays the test results of students conducted to date in the academic years 2021/2022 and 2022/2023. The Levene column shows the significance (p -value) obtained in the Levene's normality test. The t -Test column provides the p -value when comparing means, considering the result of the Levene's test. Four theoretical exams (C1, Algebra, Test1, and Test2) have been conducted in a controlled environment without access to computers or any electronic devices. The C1 exam covers complex numbers and integral calculus with applications, similar to Test1. The difference between the two exams is that the former focuses on problem-solving, while the latter emphasizes theoretical concepts with answer options that penalize in case of error. The same situation occurs for the Algebra exam and Test2. Both exams cover algebra concepts (matrices, determinants, linear equation systems, vector spaces, Euclidean spaces, linear applications, and diagonalization), but the former is centered on problem-solving, while the latter focuses on more theoretical concepts. It can be observed that there is no significant difference between the scores of C1 and Algebra, but there is in the scores of the tests. This may be due to various reasons and normal score variability; however, this difference is not evident when comparing scores from previous courses.

Next, the scores from lab sessions are compared, in which students solve problems previously prepared outside the classroom. Before each lab session, students can consult the professor with any doubts. For these exams, they have access to Mathematica, which means students can use computers. Although it is a more or less controlled environment, students could potentially access the internet since there are no restrictions on the computer connections. In approximately 50% of the laboratory sessions, there is a significant difference between the scores obtained in 2022 and those obtained in 2021, the former being higher.

Table 11. Group statistics.

Test	Year	<i>n</i>	Mean	SD	Std. Error	Levene	<i>t</i> -Test
C1	2021	124	4.8754	2.58754	0.23237	0.04	0.71
	2022	121	4.7434	2.94837	0.26803		
Algebra	2021	125	7.0921	2.35142	0.21032	0.94	0.23
	2022	121	7.4616	2.45038	0.22276		
Test1	2021	119	3.5356	1.37149	0.12572	0.17	0.00
	2022	122	1.2266	1.11864	0.10128		
Test2	2021	122	1.8521	1.43237	0.12968	0.85	0.00
	2022	121	2.9012	1.44477	0.13134		
ExPL	2021	126	7.0820	1.99573	0.17779	0.68	0.94
	2022	120	7.1005	2.06535	0.18854		
EvSes00	2021	116	8.0853	3.43482	0.31891	0.00	0.00
	2022	113	9.6018	1.68788	0.15878		
EvSes01	2021	125	8.6780	2.06102	0.18434	0.00	0.05
	2022	122	9.1511	1.71348	0.15513		
EvSes02	2021	126	9.0167	1.93803	0.17265	0.01	0.09
	2022	121	9.3674	1.32594	0.12054		
EvSes03	2021	126	8.0381	2.06083	0.18359	0.07	0.72
	2022	122	8.1254	1.79313	0.16234		
EvSes04	2021	123	8.7244	1.64499	0.14832	0.00	0.00
	2022	123	7.4617	2.24349	0.20229		
EvSes05	2021	126	9.1341	1.44364	0.12861	0.00	0.02
	2022	123	8.6431	1.94583	0.17545		
EvSes06	2021	127	9.7953	0.84833	0.07528	0.11	0.31
	2022	120	9.6813	0.90920	0.08300		
EvSes07	2021	122	9.7533	0.91283	0.08264	0.11	0.17
	2022	118	9.5805	1.01056	0.09303		
EvSes08	2021	124	8.6823	2.08358	0.18711	0.00	0.04
	2022	117	9.3744	1.59061	0.14705		
EvSes09	2021	126	7.7040	2.25148	0.20058	0.04	0.00
	2022	117	8.6530	1.77524	0.16412		
EvSes10	2021	123	8.9943	1.78736	0.16116	0.00	0.06
	2022	118	9.5381	1.20256	0.11070		
EvSes11	2021	125	8.9984	1.41632	0.12668	0.23	0.56
	2022	116	9.1135	1.66336	0.15444		
EvSes12	2021	125	8.5880	2.09710	0.18757	0.00	0.04
	2022	120	9.1033	1.69185	0.15444		
EvSes13	2021	123	8.2247	2.54037	0.22906	0.00	0.00
	2022	119	9.2714	1.71755	0.15745		
EvSes14	2021	109	9.3793	1.09188	0.10458	0.92	0.75
	2022	104	9.4288	1.19112	0.11680		
EvSes15	2021	108	7.2717	2.06008	0.19823	0.67	0.00
	2022	114	8.1136	2.19190	0.20529		
EvSes16	2021	118	8.6002	1.63740	0.15073	0.52	0.38
	2022	114	8.7895	1.64771	0.15432		
EvSes17	2021	118	9.4215	1.16241	0.10701	0.07	0.49
	2022	111	9.2986	1.50465	0.14282		
EvSes18	2021	118	9.4758	0.87958	0.08097	0.06	0.06
	2022	115	9.6757	0.70395	0.06564		

4. Discussion

B-learning methodologies are meant to promote the active participation of students, who complete tasks designed to improve and strengthen their knowledge, competencies, and skills [30,33,34,43]. Even if a correct solution is not reached, attempting to solve problems strengthens critical thinking and improves learning and deductive abilities. Repetitive and simple tasks also aim to ensure the correct assimilation of knowledge within a broad context of questions. The subject matter addressed in this article is mathematical, which requires specific skills based on the correct assimilation of content, practice, and application, as well as reinforcement through activities. In contrast to other subjects, the use of a mathematical language different from the one used conversationally implies a need for additional learning support. However, the use of ChatGPT could weaken this support if it becomes capable of solving the problems raised and explaining the calculations made. On the other hand, ChatGPT can be helpful if used properly, as it provides a detailed description of the mathematical knowledge required to solve problems.

Regarding student performance, when comparing the results to those obtained in the previous academic year, a slight increase in the scores of the practical sessions can be observed, in which the environment is not entirely controlled, as students have access to the internet. This could indicate better performance, assuming students are truthful in the survey and use ChatGPT for preparing activities. However, this increase in scores is not noticeable in exams held in more controlled environments, where no significant differences can be observed, neither for better nor for worse.

Regarding the use of GBL and dEERs to promote student motivation and competency development, the use of ChatGPT has the potential to significantly affect its usage and the information collected. The data collected from the games are used to improve the students' learning experience through a feedforward strategy [44]. However, if these data are altered, the evaluation of competencies and content is highly affected, which may prevent deficiencies shown during the game from being reinforced in the future. This could have very negative repercussions on the design of future activities.

The student's attitude will determine, as with other technologies, whether the use of ChatGPT in active methodologies will have benefits or drawbacks [18,20]. Since COVID-19 pandemic, students have an improved wide range of computer and digital tools at their disposal that facilitate learning [29], but the introduction of ChatGPT with the ability to directly customize the problem posed can greatly facilitate the search process. Consequently, the focus of attention shifts from an active search to mainly analyzing whether the answer is correct or not. If students rely solely on ChatGPT to find answers instead of training their skills, the effectiveness of FT-based methodologies could be significantly diminished. In the STEM area, the response capabilities of ChatGPT pose a risk to the integrity of the learning process if they prevent the acquisition of skills. However, the doubts generated regarding the viability and correctness of the responses generated so far can promote critical thinking.

Nevertheless, ChatGPT has also positive features for blended learning environments such as, easy access to vast information to supplement their learning resources, quick assistance with homework, assignments, or clarifying doubts, and a strong capability to adapt to users based on individual needs. In the authors' opinion, this new tool can facilitate obtaining answers and facilitate knowledge acquisition. However, its potential effects on educational development and the design of activities require evaluation, as ChatGPT's response capacity can alter the learning process [12–14]. A word cloud has been generated with recent literature about ChatGPT in education (see Figure 2).

- Teacher–student interaction: The regular use of this tool when encountering difficulties in the learning process can substantially reduce the amount of interaction between teachers and students. This reduces the teacher’s opportunities to supervise and guide the students in the assimilation of knowledge and competencies.
- Not controlled environment: Although students have access to a wealth of information on the internet, books, videos, etc., in a blended learning environment, many activities do not take place in a controlled setting. Teachers rely on students using technology to obtain certain answers. However, ChatGPT’s adaptability and ability to personalize the problems posed may oversimplify the information-seeking process, the ability to critically analyze responses, and weaken the learning process.
- Assessment challenges: The emergence of ChatGPT calls into question the usual way of assessing content acquisition. The results presented by students and the content generated by them (essays, articles, and projects) must be carefully monitored. It will take time to establish ChatGPT’s potential and determine which tests will and will not be representative of the knowledge generated and acquired.

To address these concerns, educators should strive to use ChatGPT as a supplementary tool rather than a replacement for traditional teaching methods, carefully monitor its usage, and promote critical thinking and evaluation of AI-generated content.

5. Conclusions

As AI continues to gain prominence in education, new challenges will arise in developing effective teaching methodologies that leverage the potential of these tools while addressing their limitations.

The consequences of the emergence of AI in the academic world will need to be assessed as the outcomes of implementing these tools become more measurable. In controlled environments, such as the classroom with a classic face-to-face methodology, the use of AI can be minimized, simply by restricting access to the network and mobile devices. In these same controlled settings, as students must demonstrate the acquisition of knowledge and competencies at different assessment points throughout the course, it is expected that the use of AI will be merely anecdotal, being completely inappropriate and reprehensible its use during the tests or activities.

It is crucial to proactively address these challenges to ensure that students continue to receive a high-quality education that prepares them for the demands of the future. In our opinion, the digital elements of blended learning methodologies (online exams, quizzes, knowledge reinforcement exercises, games) are the ones that carry the greatest risk of being oversimplified by AI.

Despite the recent advent of ChatGPT and risk of wrongful answers, its ability to learn and adapt is a significant advantage over other sources of information, which may also contain incorrect or outdated information. In addition, the personalization of the problems and the detailed guidance provided by ChatGPT have been highlighted by students as key strengths.

The results of this study show that students have a high level of confidence in the accuracy of ChatGPT’s answers, with a high percentage of correct responses when compared to the numerical solutions provided in the activities. Furthermore, ChatGPT not only provides solutions to the mathematical problems posed but also offers a step-by-step guide to the process required for their solution, which enhances the student’s understanding of the problem-solving process. Nevertheless, it is important to note that the use of ChatGPT may have implications for the development of critical thinking and problem-solving skills in students. Therefore, it is crucial to strike a balance between leveraging the benefits of AI and ensuring that students develop the necessary competencies to succeed in their academic and professional lives.

In conclusion, ChatGPT has both advantages and disadvantages in blended learning environments. While it offers easy access to a huge amount of information and educational assistance, it also raises concerns about the ability to assess correctly the learning progress of the students, ethical use, and oversimplification of learning process. Successful integration

of ChatGPT requires a balanced approach, where it complements human interaction and guidance. Teachers and educational institutions must carefully monitor its use to ensure it supports the learning process rather than hindering it.

Author Contributions: Conceptualization, S.M.-L., E.V.-F. and A.N.-P.; methodology, S.M.-L. and J.A.M.-F.; software, A.N.-P.; validation, L.M.S.-R., S.M.-L. and E.V.-F.; formal analysis, S.M.-L. and A.N.-P.; investigation, L.M.S.-R.; resources, A.N.-P. and E.V.-F.; data curation, S.M.-L. and J.A.M.-F.; writing—original draft preparation, S.M.-L. and L.M.S.-R.; writing—review and editing, L.M.S.-R.; visualization, E.V.-F.; supervision, L.M.S.-R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Technical University of Valencia (UPV), “Convocatoria A + D, Proyectos de Innovación Mejora Educativa, grant number PIME/21-22/284”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the authors upon reasonable request.

Acknowledgments: This research has been developed within UPV—GRIM4E (GRoup of Innovative Methodologies and assessment For engineering Education). The authors thank the anonymous reviewers for their constructive suggestions. The corresponding author is very grateful to the Applied Sciences Editorial Office, and Guest Editors Maxim Mozgovoy and Paolo Burelli, for their gentle invitation to submit this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
BL	Blended Learning
CT	Critical Thinking (skill)
DCDA	Dynamical Continuous Discrete Assessment
ER	Escape Room
EER	Educational Escape Room
dEER	Digital Educational Escape Room
FT	Flip-Teaching
GBL	Game-Based Learning
GPT	Generative Pre-trained Transformer (model)
GW	Group Work (skill)
NMS	Numerical Mathematical Solution
PS	Problem Solving (skill)
STEM	Science, Technology, Engineering, and Mathematics
SD	Standard Deviation
TMS	Theoretical Mathematical Solution

References

1. OpenAI. Available online: <https://openai.com/> (accessed on 26 March 2023).
2. Graf, A.; Bernardi, R.E. ChatGPT in Research: Balancing Ethics, Transparency and Advancement. *Neuroscience* **2023**, *515*, 71–73. [CrossRef] [PubMed]
3. Okaibedi Eke, D. ChatGPT and the rise of generative AI: Threat to academic integrity? *J. Responsible Technol.* **2023**, *13*, 100060. [CrossRef]
4. Haleem, A.; Javaid, M.; Pratap Singh, R. An era of ChatGPT as a significant futuristic support tool: A study on features, abilities, and challenges. *Bench Coun. Trans. Benchmarks Stand. Eval.* **2022**, *2*, 100089. [CrossRef]
5. Rytr. Available online: <https://rytr.me/> (accessed on 26 March 2023).
6. Jasper. Available online: <https://www.jasper.ai> (accessed on 26 March 2023).
7. CopyAI. Available online: <https://www.copy.ai/?via=start> (accessed on 26 March 2023).
8. Writesonic. Available online: <https://writesonic.com/?via=sign-up-now> (accessed on 26 March 2023).

9. Kafkai. Available online: <https://kafkai.com> (accessed on 26 March 2023).
10. Copysmith. Available online: <https://app.copysmith.ai> (accessed on 26 March 2023).
11. Article Forge. Available online: <https://www.articleforge.com/> (accessed on 26 March 2023).
12. Klang, E.; Levy-Mendelovich, S. Evaluation of OpenAI's large language model as a new tool for writing papers in the field of thrombosis and hemostasis. *J. Thromb. Haemost.* **2023**, *in press*. [[CrossRef](#)] [[PubMed](#)]
13. Gilat, R.; Cole, B.J. How Will Artificial Intelligence Affect Scientific Writing, Reviewing and Editing? The Future is Here.... *Arthrosc. J. Arthrosc. Relat. Surg.* **2023**, *in press*. [[CrossRef](#)] [[PubMed](#)]
14. Alser, M.; Waisberg, E. Concerns with the usage of ChatGPT in Academia and Medicine: A viewpoint. *Am. J. Med. Open* **2023**, *100036*. [[CrossRef](#)]
15. Gilson, A.; Safranek, C.W.; Huang, T.; Socrates, V.; Chi, L.; Taylor, R.A.; Chartash, D. How Does ChatGPT Perform on the United States Medical Licensing Examination? The Implications of Large Language Models for Medical Education and Knowledge Assessment. *JMIR Med. Educ.* **2023**, *9*, e45312. [[CrossRef](#)]
16. Rudolph, J.; Tan, S.; Tan, S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *J. Appl. Learn. Teach.* **2023**, *6*, 1–22. [[CrossRef](#)]
17. Kung, T.H.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G.; Maningo, J.; et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLoS Digit. Health* **2023**, *2*, e0000198. [[CrossRef](#)]
18. Shen, Y.; Heacock, L.; Elias, J.; Hentel, K.D.; Reig, B.; Shih, G.; Moy L. ChatGPT and other large language models are double-edged swords. *Radiology* **2023**, *in press*. [[CrossRef](#)]
19. Chen, T.J. ChatGPT and other artificial intelligence applications speed up scientific writing. *J. Chin. Med. Assoc.* **2023**, *in press*. [[CrossRef](#)] [[PubMed](#)]
20. Mhlanga, D. Open AI in Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning. *SSRN* **2023**. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354422 (accessed on 5 May 2023). [[CrossRef](#)]
21. Sallam, M. ChatGPT Utility in Healthcare Education, Research, and Practice: Systematic Review on the Promising Perspectives and Valid Concerns. *Healthcare* **2023**, *11*, 887. [[CrossRef](#)]
22. Pal, S. Performing Effective Research Using ChatGPT. *Indian J. Comput. Sci.* **2022**, *7*, 1–10. [[CrossRef](#)]
23. Lim, W.M.; Gunasekara, A.; Pallant, J.L.; Pallant, J.I.; Pechenkina, E. Generative AI and the future of education: Ragnarök or reformation? A paradoxical perspective from management educators. *Int. J. Manag. Educ.* **2023**, *21*, 100790. [[CrossRef](#)]
24. European Ministers of Education. Available online: www.ehea.info/cid100210/ministerial-conference-bologna-1999.htm (accessed on 2 May 2023).
25. Chen, L.; Tang, X.-J.; Liu, Q.; Zhang, X. Self-directed learning: Alternative for traditional classroom learning in undergraduate ophthalmic education during the COVID-19 pandemic in China. *Heliyon* **2023**, *e15632*. [[CrossRef](#)]
26. Isha, S.; Wibawarta, B. The impact of the COVID-19 pandemic on elementary school education in Japan. *Int. J. Educ. Res. Open* **2023**, *4*, 100239. [[CrossRef](#)]
27. Martin, B.; Kaminski-Ozturk, N.; Smiley, R.; Spector, N.; Silvestre, J.; Bowles, W.; Alexander, M. Assessing the Impact of the COVID-19 Pandemic on Nursing Education: A National Study of Prelicensure RN Programs. *J. Nurs. Regul.* **2023**, *14*, S1–S67. [[CrossRef](#)]
28. Hoofman, J.; Secord, E. The Effect of COVID-19 on Education. *Pediatr. Clin. N. Am.* **2021**, *68*, 1071–1079. [[CrossRef](#)] [[PubMed](#)]
29. Sánchez Ruiz, L.M.; Moll-López, S.; Morano-Fernández, J.A.; Llobregat-Gómez, N. B-Learning and Technology: Enablers for University Education Resilience. An Experience Case under COVID-19 in Spain. *Sustainability* **2021**, *13*, 3532. [[CrossRef](#)]
30. Müller, C.; Mildenerberger, T. Facilitating flexible learning by replacing classroom time with an online learning environment: A systematic review of blended learning in higher education. *Educ. Res. Rev.* **2021**, *34*, 100394. [[CrossRef](#)]
31. Du, L.; Zhao, L.; Xu, T.; Wang, Y.; Zu, W.; Huang, X.; Nie, W.; Wang, L. Blended learning vs. traditional teaching: The potential of a novel teaching strategy in nursing education—A systematic review and meta-analysis. *Nurse Educ. Pract.* **2022**, *63*, 103354. [[CrossRef](#)]
32. Boelens, R.; De Wever, B.; Voet, M. Four key challenges to the design of blended learning: A systematic literature review. *Educ. Res. Rev.* **2017**, *22*, 1–18. [[CrossRef](#)]
33. Hrastinski, S. What do we mean by blended learning? *TechTrends* **2019**, *63*, 564–569. [[CrossRef](#)]
34. Saichaie, K. Blended, flipped, and hybrid learning. *Defin. Dev. Dir.* **2020**, *164*, 95–104. [[CrossRef](#)]
35. Harrison, D.J.; Saito, L.; Markee, N.; Herzog, S. Assessing the effectiveness of a hybrid-flipped model of learning on fluid mechanics instruction: Overall course performance, homework, and far- and near-transfer of learning. *Eur. J. Eng. Educ.* **2017**, *42*, 712–728. [[CrossRef](#)]
36. Andrade, M.S.; Alden-Rivers, B. Developing a framework for sustainable growth of flexible learning opportunities. *High. Educ. Pedagog.* **2019**, *4*, 1–16. [[CrossRef](#)]
37. Asarta, C.J.; Schmidt, J.R. The choice of reduced seat time in a blended course. *Internet High. Educ.* **2015**, *27*, 24–31. [[CrossRef](#)]
38. Ashby, J.; Sadara, W.A.; McNary, S.W. Comparing student success between developmental math courses offered online, blended, and face-to-face. *J. Interact. Online Learn.* **2011**, *10*, 128–140.
39. Baeppler, P.; Walker, J.D.; Driessen, M. It's not about seat time: Blending, flipping, and efficiency in active learning classrooms. *Comput. Educ.* **2014**, *78*, 227–236. [[CrossRef](#)]

40. Bernard, R.M.; Borokhovski, E.; Schmid, R.F.; Tamim, R.M.; Abrami, P.C. A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *J. Comput. High. Educ.* **2014**, *26*, 87–122. [\[CrossRef\]](#)
41. Dziuban, C.; Graham, C.R.; Moskal, P.D.; Norberg, A.; Sicilia, N. Blended learning: The new normal and emerging technologies. *Int. J. Educ. Technol. High. Educ.* **2018**, *15*, 1–16. [\[CrossRef\]](#)
42. Gagnon, M.P.; Gagnon, J.; Desmartis, M.; Njoya, M. The impact of blended teaching on knowledge, satisfaction, and self-directed learning in nursing undergraduates: A randomized, controlled trial. *Nurs. Educ. Perspect.* **2013**, *34*, 377–382. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Melton, B.F.; Bland, H.; Chopak-Foss, J. Achievement and satisfaction in blended learning versus traditional general health course designs. *Int. J. Scholarsh. Teach. Learn.* **2009**, *3*, 26. [\[CrossRef\]](#)
44. Sánchez-Ruiz, L.M.; Moll-López, S.; Morano-Fernández, J.A.; Roselló, D. Dynamical Continuous Discrete Assessment of Competencies Achievement: An Approach to Continuous Assessment. *Mathematics* **2021**, *9*, 2082. [\[CrossRef\]](#)
45. Alenezi, A.; Karim, A.; Veloo, A. An empirical investigation into the role of enjoyment, computer anxiety, computer self-efficacy and Internet experience in influencing the students' intention to use e-learning: A case study from Saudi Arabian governmental universities. *Turk. Online J. Educ. Technol.* **2010**, *9*, 22–34.
46. Ocak, M.A. Why are faculty members not teaching blended courses? Insights from faculty members. *Comput. Educ.* **2011**, *56*, 689–699. [\[CrossRef\]](#)
47. So, H.J.; Brush, T.A. Student perceptions of collaborative learning, social presence, and satisfaction in a blended learning environment: Relationships and critical factors. *Comput. Educ.* **2008**, *51*, 318–336. [\[CrossRef\]](#)
48. Aycock, A.; Garnham, C.; Kaleta, R. Lessons learned from the hybrid course project. *Teach. Technol. Today* **2002**, *8*, 1–5.
49. Kaplan, A.; Ozdemir, C.; Kaplan, O. The Effect of the Flipped Classroom Model on Teaching Clinical Practice Skills. *J. Emerg. Nurs.* **2022**, *49*, 124–133. [\[CrossRef\]](#)
50. Joy, P.; Panwar, R.; Adibatti, M. Flipped classroom – A student perspective of an innovative teaching method during the times of pandemic. *Educ. Méd.* **2023**, *24*, 10079017. [\[CrossRef\]](#)
51. Barranquero-Herbosa, M.; Abajas-Bustillo, R.; Ortego-Maté, C. Effectiveness of flipped classroom in nursing education: A systematic review of systematic and integrative reviews. *Int. J. Nurs. Stud.* **2022**, *135*, 104327. [\[CrossRef\]](#) [\[PubMed\]](#)
52. Cortese, G.; Greif, R.; Charco Mora, P. Flipped classroom and a new hybrid “Teach the Airway Teacher” course: An innovative development in airway teaching? *Trends Anaesth. Crit. Care* **2022**, *42*, 1–3. [\[CrossRef\]](#)
53. Connolly, T.M.; Boyle, E.A.; MacArthur, E.; Hainey, T.; Boyle, J.M. A systematic literature review of empirical evidence on computer games and serious games. *Comput. Educ.* **2012**, *59*, 661–686. [\[CrossRef\]](#)
54. Boyle, E.A.; Hainey, T.; Connolly, T.M.; Gray, G.; Earp, J.; Ott, M.; Lim, T.; Ninaus, M.; Ribeiro, C.; Pereira, J. An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games. *Comput. Educ.* **2016**, *94*, 178–192. [\[CrossRef\]](#)
55. Hainey, T.; Connolly, T.M.; Boyle, E.A.; Wilson, A.; Razak, A. A systematic literature review of games-based learning empirical evidence in primary education. *Comput. Educ.* **2016**, *102*, 202–223. [\[CrossRef\]](#)
56. Bybee, R.W. *The Case for STEM Education: Challenges and Opportunities*; NSTA Press: Arlington, VA, USA, 2013; pp. 1–116.
57. Mellado, V.; Borrachero, A.B.; Brígido, M.; Melo, L.V.; Davila, M.A.; Canada, F.; Conde, M.C.; Costillo, E.; Esteban, R.; Martínez, G.; et al. Las emociones en la enseñanza de las ciencias/Emotions in science teaching. *Enseñanza Cienc.* **2014**, *32*, 11–36.
58. Ebner, M.; Holzinger, A. Successful implementation of user-centered game based learning in higher education: An example from civil engineering. *Comput. Educ.* **2007**, *49*, 873–890. [\[CrossRef\]](#)
59. Menon, D.; Romero, M. Game mechanics supporting a learning and playful experience in educational escape games. In *Global Perspectives on Gameful and Playful Teaching and Learning*; IGI Global: Hershey, PA, USA, 2020; pp. 143–162.
60. Zamora-Polo, F.; Corrales-Serrano, M.; Sanchez-Martín, J.; Espejo-Antúnez, L. Nonscientific university students training in general science using an active-learning merged pedagogy: Gamification in a flipped classroom. *Educ. Sci.* **2019**, *9*, 297–315. [\[CrossRef\]](#)
61. Ross, R.; Bell, C. Turning the classroom into an escape room with decoder hardware to increase student engagement. In Proceedings of the 2019 IEEE Conference on Games (CoG), London, UK, 20–23 August 2019; pp. 1–4.
62. Sanchez-Martín, J.; Corrales-Serrano, M.; Luque-Sendra, A.; Zamora-Polo, F. Exit for success. Gamifying science and technology for university students using escape-room. A preliminary approach. *Heliyon* **2020**, *6*, e04340. [\[CrossRef\]](#)
63. Lopez-Pernas, S.; Gordillo, A.; Barra, E.; Quemada, J. Examining the Use of an Educational Escape Room for Teaching Programming in a Higher Education Setting. *IEEE Access* **2019**, *7*, 31723–31737. [\[CrossRef\]](#)
64. Charlo, J.C.P. Educational Escape Rooms as a Tool for Horizontal Mathematization: Learning Process Evidence. *Educ. Sci.* **2020**, *10*, 213. [\[CrossRef\]](#)
65. Gordillo, A.; López-Fernández, D.; López-Pernas, S.; Quemada, J. Evaluating an Educational Escape Room Conducted Remotely for Teaching Software Engineering. *IEEE Access* **2020**, *8*, 225032–225051. [\[CrossRef\]](#)
66. Zhang, F.; Doroudian, A.; Kaufman, D.; Hausknecht, S.; Jeremic, J.; Owens, H. Employing a user-centered design process to create a multiplayer online escape game for older adults. In *Human Aspects of IT for the Aged Population. Applications, Services and Contexts, Proceedings of the Third International Conference, ITAP 2017, Vancouver, BC, Canada, 9–14 July 2017*; Springer: Cham, Switzerland, 2017; pp. 296–307.

67. Nicholson, S. Ask why: Creating a better player experience through environmental storytelling and consistency in escape room design. *Meaningful Play* **2016**, 521–556. Available online: <https://scottnicholson.com/pubs/askwhy.pdf> (accessed on 5 May 2023).
68. Bassford, M.L.; Crisp, A.; O’Sullivan, A.; Bacon, J.; Fowler, M. CrashEd—A live immersive, learning experience embedding STEM subjects in a realistic, interactive crime scene. *Res. Learn. Technol.* **2016**, *24*, 30089–30093. [CrossRef]
69. Sánchez-Ruiz, L.M.; López-Alfonso, S.; Moll-López, S.; Moraño-Fernández, J.A.; Vega-Fleitas, E. Educational Digital Escape Rooms Footprint on Students’ Feelings: A Case Study within Aerospace Engineering. *Information* **2022**, *13*, 478. [CrossRef]
70. Likert, R. A Technique for the Measurement of Attitudes. *Arch. Psychol.* **1932**, *140*, 1–55.
71. Typeform. Available online: <https://www.typeform.com> (accessed on 26 March 2023).
72. Shapiro, S.S.; Wilk, M.B. An analysis of variance test for normality (complete samples). *Biometrika* **1965**, *52*, 591–611. [CrossRef]
73. RPG Maker MZ Software. Available online: <https://www.rpgmakerweb.com/products/rpg-maker-mz> (accessed on 4 April 2023).
74. Wolfram Mathematica. Available online: <https://www.wolfram.com/mathematica/> (accessed on 4 April 2023).
75. Wolfram Alpha. Available online: <https://www.wolframalpha.com/> (accessed on 4 April 2023).
76. Nube de Palabras. Available online: <https://www.nubedepalabras.es/> (accessed on 5 May 2023).
77. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* **2023**, *13*, 410. [CrossRef]
78. Frieder, S.; Pinchetti, L.; Griffiths, R.R.; Salvatori, T.; Lukasiewicz, T.; Petersen, P.C.; Chevalier, A.; Berner, J. Mathematical Capabilities of ChatGPT. *arXiv* **2023**, arXiv:2301.13867.
79. Geerling, W.; Mateer, G.D.; Wooten, J.; Damodaran, N. Is ChatGPT Smarter than a Student in Principles of Economics? *SSRN* **2023**, 4356034. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4356034 (accessed on 5 May 2023). [CrossRef]
80. Mogali, S.R. Initial Impressions of ChatGPT for Anatomy Education. *Anat. Sci. Educ.* **2023**, *in press*. [CrossRef]
81. Newton, P.M. ChatGPT Performance on MCQ-based Exams. *EdArXiv* **2023**. [CrossRef]
82. Jalil, S.; Rafi, S.; LaToza, T.D.; Moran, K.; Lam, W. ChatGPT and Software Testing Education: Promises & Perils. *arXiv* **2023**, arXiv:2302.03287.
83. King, M.R. A Conversation on Artificial Intelligence, Chatbots, and Plagiarism in Higher Education. *Cell. Mol. Bioeng.* **2023**, *16*, 1–2. [CrossRef]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.