Article

Not peer-reviewed version

Empowering Education 4.0: Impact of Using ChatGPT as a Virtual Mentor on K-12 Students Learning Science

Rafael Castañeda *, Laura Mercadé, Víctor J. Gómez, Teresa Mengual, Francisco Javier Díaz-Fernández, Miguel Sinusia Lozano, Juan Navarro Arenas, Ángela Barreda, Maribel Gómez-Gómez, Elena Pinilla Cienfuegos, David Ortiz de Zárate *

Posted Date: 17 September 2024

doi: 10.20944/preprints202409.1323.v1

Keywords: education 4.0; artificial intelligence; blended learning; ChatGPT; virtual mentor; K-12; science education

Article

Empowering Education 4.0: Impact of Using ChatGPT as a Virtual Mentor on K-12 Students Learning Science

Rafael Castañeda [1,*], Laura Mercadé [2], Víctor Jesús Gómez [2], Teresa Mengual [2], Francisco Javier Díaz-Fernández [2], Miguel Sinusia Lozano [2], Juan Navarro Arenas [3], Ángela Barreda [4], Maribel Gómez [2], Elena Pinilla-Cienfuegos [2] and David Ortiz de Zárate [2,*]

[1] IES de Benaguasil, Calle Segorbe 2, 46180 Benaguasil, València, Spain
[2] Nanophotonics Technology Center (NTC), Universitat Politècnica de València (UPV), Camí de Vera s/n, 46022 València, Spain
[3] Center for Nanotechnology (CeNTech), Raum CenTech II E.15, Heisenbergstraße 11, 48149 Münster, Germany
[4] Group of Displays and Photonics Applications, Carlos III University of Madrid, Avda. de la Universidad, 30, Leganés, 28911 Madrid, Spain
* Correspondence: r.castanedasanchez@edu.gva.es or racassan@gmail.com (R.C.); daorde@ntc.upv.es (D.O.d.Z.)

Abstract: Education 4.0 arises to provide citizens with the technical/digital competencies and cognitive/interpersonal skills demanded by Industry 4.0. New technologies drive this change, though time-independent learning remains a challenge, because students might face a lack of support, advice and surveillance when teachers are unavailable. This study proposes complementing presential lessons with online learning driven by ChatGPT, applied as an educational tool able to mentor K-12 students learning science at home. First, ChatGPT's performance in the field of K-12 science is evaluated, scoring A (9.3/10 in 2023, and 9.7/10 in 2024) and providing detailed, analytic, meaningful, and human-like answers. Then, an empirical interventional study is performed to assess the impact of using ChatGPT as a virtual mentor on real K-12 students. After the intervention, the grades of students in the experimental group improved 30%, and 70% of students stated a positive perception of the AI, suggesting a positive impact of the proposed educational approach. After discussion, the study concludes ChatGPT might be a useful educational tool able to provide K-12 students learning science with the functional and social/emotional support they might require, democratizing a higher level of knowledge acquisition and promoting students' autonomy, security and self-efficacy. The results probe ChatGPT's remarkable capacity (and immense potential) to assist teachers in their mentoring tasks, laying the foundations of virtual mentoring and paving the way for future research aimed at obtaining a more realistic view of the AI impact on education.

Keywords: education 4.0; artificial intelligence; blended learning; ChatGPT; virtual mentor; K-12; science education

1. Introduction

"It's the end of the world as we know it" is not only the title of a song, but also a proper description of current societies worldwide, facing the deep, and quick changes impelled by the Fourth Industrial Revolution (4IR) that every citizen is witnessing [1,2].

The previous industrial revolutions were driven by technologic advancements, and their effect on the industry and societies was profound [3]. While the 1IR employed steam and waterpower to mechanize manufacturing, the 2IR enabled mass production through the use of electricity and the division of labor, and the 3IR was powered by electronics and computerization, allowing the

 doi:10.20944/preprints202409.1323.v1

automation of manufacturing and the analog to digital transition, changing the world's capacity to store information in digital format from less than 1% in 1980s, to more than 99% by 2014 [4].

Conversely, the present industrial revolution (4IR) pursues a new paradigm of smart, autonomous, and sustainable manufacturing, the so-called Industry 4.0 [5], whose foundational pillars were conceived to empower every citizen and every government to build a better and more inclusive, human-centered world [3]. This revolution still exploits the key developments to digitalize the information set by the Digital Revolution (3IR) as early as 1947 [6], that is transistors, integrated circuits, microprocessors and computers, Internet, digital mobile phones and even digital TV. However, the 4IR is mainly driven by a set of disruptive technologies that blur the lines between the physical, digital, and biological worlds (through cyber-physical systems) [3]. These technologies aim at [5,7]:

(1) *Increasing connectivity, data and computational power* (cloud technology, smart sensors and actuators -even wearables-, blockchain…)
(2) *Boosting analytics and system intelligence* (advanced analytics, machine learning, neural networks, artificial intelligence -AI-…)
(3) *Promoting machine-machine and human-machine interaction* (extended reality, XR -including virtual, augmented and mixed reality, that is VR, AR and MR, respectively-, digital twins, robotics, automation, autonomous guided vehicles, Internet of Things, Internet of Systems…)
(4) *Enhancing advanced engineering* (additive manufacturing such as 3D printing, ICTs, nanotechnology, renewable energies, biotechnology…)

All these disruptive technologies are leading Industry 4.0 towards the concept of Smart Manufacturing (as production becomes faster, closer and more responsive to customer/market requirements), exhibiting unprecedented degrees of interoperability, virtualization, decentralization, real-time capability, modularity, information transparency and technical assistance [2,8]. Consequently, the 4IR is not only improving the efficiency of business and keeping billions of people interconnected, but also enhancing sustainability through better management of resources, contributing to regenerating the natural environment and potentially reverting the damage earlier industrial revolutions provoked [9].

Thus, the 4IR is already changing the way we perceive the world, work, think, relate and definitely live in an unprecedented scale, scope and complexity. Anyway, the opportunity and the responsibility to build a better and sustainable future through the Industry 4.0 must be embraced by worldwide governments, companies, industries, academia and civil society [3].

Since technology is only one part of the equation, workers and citizens must adapt their skills to a new future where machines will take care of heavy duties, and humans will just attend to derived problems. Therefore, all the abilities that machines lack will now be very appreciated in the labor market. These competencies can be divided into three categories [10]:

(1) *High-level technical skills*: knowledge in ICT - information and communication technologies - sciences, Big Data and data analysis, network management, programming, 3D printing, nano/biotechnology…

(2) *High-order cognitive skills*: critical thinking, problem-solving, decision-making abilities…

(3) *Human and interpersonal skills*: creativity, social, emotional capabilities…

Societies might empower their citizens with those skills and digital competencies, and the best way is education. Old-school education based on the *Empty Container Paradigm* (students are an empty container that must be filled with knowledge), presented several problems to satisfy the fit-for-purpose criterium within the context of Industry 4.0 and the 4IR [11,12]:

(1) *The current teaching process has plenty of room for interactivity improvement* (which is required for a better and longer learning)

(2) *Assessments only evaluate the amount of learned knowledge*, but not the acquired competencies/skills

(3) *Wide time gap between receiving knowledge and its application in practice*.

Therefore, the education evolved over the years until reaching the concept of Education 4.0 [13–16], aimed at providing students with not only technical/digital competencies, but also cognitive and

 doi:10.20944/preprints202409.1323.v1

human/interpersonal skills required by the society and the Industry 4.0. Indeed, it was clearly boosted worldwide by COVID-19-related school closures [17–20]. The foundations of Education 4.0 might be summarized as [11,16]:

(1) *New student-centered learning strategies (heutagogy, peeragogy, cybergogy…)*
(2) *Location and time-independent learning*
(3) *Personalized learning*
(4) *Interactive/collaborative learning*
(5) *Gamification to raise engagement*
(6) *Online sources of information* (web, massive open online courses -MOOCs-…)
(7) *Teacher to mentor transition*

Many of them are currently being fulfilled by using 4IR technologies. According to bibliography [17,21,22], the most frequently applied technologies within Education 4.0 are VR, AR, MR, eye-tracking, learning-analytics, robotics and simulation, besides the tools enabling online learning such as streaming lectures, virtual classrooms, digital boards, cloud systems and MOOCs. AI is also applied in combination with learning-analytics to assess the students' progress in order to find weaknesses and adapt the education process to their particular needs, enabling a more personalized education [17,23–26].

A simple problem hampering efficient time-independent learning (part of the second pillar of Education 4.0) is the availability of mentors to assist students, which cannot be complete. Therefore, there is a growing interest in MOOCs and virtual mentoring, as they are effective and complementary tools outside schools. However, MOOCs lack interactivity, mentoring capacity and freedom to provide personalized answers to the student's particular doubts, while current virtual mentoring still refers to remote mentoring, that is connecting students with their human mentors, which still does not address the availability limitation.

The use of AI in education has been explored since 1970s, when computers were devised not only as tools but also as potential tutors [27], developing a new field of research using computers to smartly coach students termed Intelligent Computer Assisted Instruction or Intelligent Tutoring Systems (ITS) [28–31]. Computer learning has evolved over time, integrating artificial intelligence, and thus, plenty dialogue-based systems have been developed [29–31] for assisting students in learning different subjects, from STEM sciences to politics, claiming to provide students with a meaningful interaction, which is key for the long-term learning [32], and pointing out student's difficulties in order to duly address them, promoting customized learning. However, ITS effectiveness has been debated over the years [33,34]. Besides, these solutions have been purposely designed for teaching, so they are forced to follow a pre-programmed method to teach students, which limits the ITS capacity of solving the student's particular doubts arising when doing homework without any tutor. This might explain why these solutions have barely been employed in experimental lessons including problem-solving and/or decision-making in chemistry, physics, and clinical fields [30].

Other researchers have assessed the use of old chatbots as "*real-time educational assistants for the teacher*", but they actually lack interactivity, just relating to students through pre-programmed answers lacking intelligence, and their only aim is raising engagement and generating statistics of student's understanding level [32,35–37]. Again, these systems lack freedom to answer the student's particular doubts.

The growth of AI during the last years has recently been quantified, not only in numbers but also in real applications within the market: the adoption of AI in companies (50%) has doubled since 2017 (20%) and the average number of AI capabilities they exploit (such as natural-language generation or computer vision), has also doubled from 1.9 in 2018 to 3.8 in 2022 [38], according to a McKinsey Global Survey [39]. In 2023, 149 new AI foundation models were released (mainly from the industry), more than double the amount appeared in 2022, and the number of AI patents sharply increased worldwide (67% from 2021 to 2022). Surprisingly, while the US is the leading source of those models, China led global AI patent origins with 61% (being the US the origin of 21%) [40]. Today, AI can be applied not only to process information and take decisions (even in fields such as

 doi:10.20944/preprints202409.1323.v1

weather forecast [41] or chemistry [42,43], among many others), but also to generate original text, images and even video content, and even if the AI has already overcome human performance on several benchmarks and domains (such as image classification, visual reasoning, and English understanding), human beings still perform better than the AI within more complex tasks like competition-level mathematics, visual commonsense reasoning and planning) [40].

Concerning natural language processing, Open AI (San Francisco, California) developed in 2018 an autoregressive language model called GPT (Generative Pre-trained Transformer), exploiting deep learning to produce human-like text from image/text inputs [44], whose last version (GPT-4, released on March 14, 2023) feeds an artificial intelligence chatbot called ChatGPT, able to write and debug computer programs [45,46], generate documents such as essays [47]; answer test questions [48], write poetry, lyrics and compose music [49], and even provide assistance in complex scientific tasks [50–52]… Despite this impressive demonstration of capabilities within a short time of existence, ChatGPT is just a large-scale multimodal model powered by AI, so its application in fields different from the strict generation of text is currently unveiling a problem related to a lack of appropriate training, guarantee of intellectual property rights preservation, and hallucinations (mistaken or nonsensical answers that seem semantically or syntactically correct), which are allegedly limited in the last version of the language model, and will tend to reduce with an increased training [53]. As a consequence, this technology is actually earning as many positive judgements as widespread criticism from artists, ethicists, academics, educators, and journalists [54–57].

There is a recent and growing interest into assessing the potential of AI-powered chatbots (such as ChatGPT) as an educational tool in quite different fields such as language, programming, mathematics, medicine and economy, among many others [58–61]. Many of those studies focus on evaluating the chatbot ability to ask and also answer particular fact-based and test questions. Even OpenAI has evaluated its own technology by means of exams purposely designed for humans. While GPT-3.5 performance lied at bottom 10% of test takers, GPT-4 outscored better than the majority of human test takers (top 10%). Those previous studies have promoted an open debate around the potential benefits and limitations of ChatGPT in the field of education [44,58,62–67], even suggesting its potential use as teaching assistant [64–66,68,69] automatically tackling duties such as creating assessment tests, grading, guiding, and recommending learning materials, though also claiming that ChatGPT "lacks the ability to provide emotional support, and facilitate critical thinking and problem-solving skills in science learning", from a theoretical perspective [68].

Concerning the use of ChatGPT in the field of science education, the studies usually exploring the generative artificial intelligence capabilities to answer a few theoretical questions [68,70,71], or applied questions such as acid/base problems [72]. Some studies have even developed a theoretical framework for applying generative AI in the field of education [73]. Nevertheless, those publications claim the future research should focus on evaluating the impact of ChatGPT (or other AI applications) on real students' learning outcomes, such as academic achievement, motivation and engagement, in different contexts, as there is still a need for real cases of AI impact on real students, that is, empirical and systematic evaluations of the use of ChatGPT on real students learning science [56,73].

Therefore, the present exploratory study will try to provide more insight in this field, by addressing the following research questions (RQs), within the frame of applying generative AI in K-12 science education:

RQ1: Does ChatGPT provide a trustworthy time-independent learning experience to K-12 students, when teachers are unavailable?

RQ2: Can ChatGPT create meaningful interactions with K-12 students?

RQ3: What is the real impact of using ChatGPT as a virtual mentor on K-12 students learning science when teachers are unavailable?Following previous works on the use of AI within an educational context [74–76], this exploratory study will address these RQs by the evaluation of ChatGPT's competence to become an educational tool aimed at providing K-12 students with a personalized, meaningful, and location- and time-independent learning, in a safe environment and real time, assisting teachers in the task of mentoring students through specific duties such as homework correcting and solving doubts at home. A special focus will be set on assessing: (a) student's

 doi:10.20944/preprints202409.1323.v1

proficiency before and after the intervention, and (b) students' perception of the AI as a useful educational tool, once duly evaluated. to the best of our knowledge, this is the first empirical assessment of the real impact of using ChatGPT as a virtual mentor on K-12 students learning chemistry and physics, within the frame of a blended-learning pedagogical approach combining constructivist/connectivist presential learning (Education 2.0 and 3.0) with student-centered self-regulated cybergogy (Education 4.0) [16].

2. Materials and Methods

The study was designed following the previously described IDEE theoretical framework for using any generative AI in the field of science education [73]. According to this, the main pillars of the study have been identified:

- *Desired outcomes*: This empirical study aims to systematically assess the real effect, possibilities and challenges of applying a complementary and well-defined use of ChatGPT outside the traditional school environment (mainly focused on correcting specific homework assignments designed by the teacher and solving students' particular doubts and needs) on K-12 (15-16-years-old) students learning chemistry and physics. This will allow finding the answers to the previous RQs, which will provide more insight regarding the use of advanced AI tools such as ChatGPT as teaching assistants in the field of science education. The outcomes that will be monitored to assess the impact of the AI on students will be their proficiency (through grades evolution) and their perception on the AI as an educational tool, before and after the intervention.
- *Appropriate level of automation*: The study has been designed within a blended-learning pedagogical approach, where the teacher role is essential as not only mentor but also facilitator [77]. Thus, K-12 students kept the constructivist/connectivist presential learning at school in combination with online learning experiences designed by the teacher (flipped-learning [78]). The only difference arose for those students in the experimental group, who might complement their homework tasks by means of ChatGPT, employed as an educational tool able to correct assignments, solve doubts and guide the students towards a better understanding of the lesson and a stronger and longer-term settlement of knowledge. Therefore, only a partial automation is considered.
- *Ethical considerations*: All procedures performed in this study, involving human participants, were in accordance with the national and European ethical standards (European Network of Research Ethics Committees), the 1964 Helsinki Declaration and its later amendments, the 1978 Belmont report, the EU Charter of Fundamental Rights (26/10/2012), and the EU General Data Protection Regulation (2016/679). As the study involved 15-16-years-old students, parental informed consent was obtained from all individual participants included in the study. Main ethical concerns discussed in bibliography are related to intellectual property, privacy, biases, fairness, accuracy, transparency, lack of robustness against "jailbreaking prompts", and the electricity and water consumption to sustain the AI servers [79–82]. In this study, the planned use of ChatGPT leaves little room for intellectual property, privacy or transparency issues. Besides, jailbreaking prompts seem not to be useful for students in this case. However, students misusing ChatGPT to do their homework instead of positively exploiting the AI to correct their homework and solve their doubts [56] might be a potential problem, but this technology is so new and attractive that students will easily be engaged to test ChatGPT and its potential benefits. Anyhow, the potential misuse might easily be detected by comparing students' grades before and after the intervention, as grades of students misusing the AI would never show any improvement. Another potential consideration might be the generation of incorrect or biased information, as the AI answers are limited by the previous training and some mathematical hallucinations have already been detected [83]. Thus, a previous validation of ChatGPT's performance in the specific field of K-12 chemistry and physics will be assessed. In the case of large language models, bias can be defined as the appearance of systematic misrepresentations, attribution errors or factual distortions based on learned patterns that might drive to supporting certain groups or ideas over different ones, preserving stereotypes or even make incorrect assumptions [84]. Training data, algorithms and other factors might contribute to the rise of demographic, cultural, linguistic, temporal, confirmation, and ideological/political biases [85]. However, these potential preexisting biases

 doi:10.20944/preprints202409.1323.v1

within the model should not affect the utility of the AI within the field of interest (K-12 science education), even if users should and will be aware of this possibility. Besides those considerations, the foreseen impact of this study on learners focuses on achieving a better understanding of the lesson, a stronger and longer-term settlement of knowledge. Concerning teachers, they would be assisted in a time- and location-independent manner by the AI in their task of mentoring students, leaving teachers more time to personally satisfy particular students' needs.

- *Evaluation of the effectiveness*: According to bibliography, the gold standard for measuring change after any intervention (i.e. within educational research) is the experimental design model [85]. In this case, the study assessed the effectiveness of the proposed educational approach through a quasi-experimental analysis, that is an empirical interventional study avoiding randomization able to determine the causal effects of an intervention (the impact of a chatbot powered by AI used as a virtual mentor on K-12 students learning chemistry and physics when their teachers are unavailable) on the target population. Randomization was not an option for the present study, as there was an interest in counting on two groups of students (the one interacting with the chatbot -experimental group- and that without any interaction with the AI -the control group-, balancing students' level of proficiency (low, medium and high), thus avoiding potential biases coming from hypothetically unbalanced groups. First, the real performance of ChatGPT in the field of chemistry and physics for K-12 students (precisely 15-16-years-old students) was systematically evaluated by the authors. The AI-powered chatbot answered a test specifically designed for real K-12 students, including a set of 52 selected theoretical questions and problems summarizing the knowledge and problem-solving skills to be acquired during a complete academic course, in a similar way to previous studies [48,59,60], always keeping in mind that this technology is not purposely designed for education, despite its great potential. No difficult nor impossible questions were removed from the set of questions as other studies did (i.e. questions demanding drawings as outputs, or analyzing images as inputs) [86], in order to obtain a fair and accurate perception of the performance of ChatGPT within this particular field, including all type of knowledge and skills requested for 15-16-years-old students learning chemistry and physics. Eleven teachers including chemists, physicists, and engineers evaluated the answers. The AI replies to theoretical questions were assessed looking for clarity, accuracy and soundness, while more applied questions such as problems were not only evaluated by the accuracy of the final result, but also by the validity and clarity of the procedure to reach that result, paying special attention to those resources enabling a stronger and longer-term knowledge settlement in a pedagogical manner. Once the theoretical performance of the chatbot in the field of interest was assessed, the authors judged the experimental capacity of this tool to assist teachers in the task of mentoring real 15-16-years-old students learning chemistry and physics when educators were unavailable, precisely in duties such as solving theoretical doubts and correcting homework assignments (including problem-solving questions) in real time and without time restrictions. Therefore, this study empirically assessed the impact of providing students with a meaningful interaction with the chatbot through which they could experience a completely personalized learning, improving their knowledge and skills while boosting their engagement. All of this could be monitored through two indicators chosen to measure the impact of ChatGPT on K-12 students learning chemistry and physics, before and after the intervention: Students' grades (taking into account both proficiency and problem-solving skills) and their perception on the AI as a useful educational tool.

Finally, among the different chatbots powered by AI (both free and paid), ChatGPT was selected to perform the experiments described within this study because of two main reasons: ChatGPT was totally free at the time (which could contribute to reducing inequities in the field of education, even if the latest version was not free for some time, and is now free again), and it exploits the original *OpenAI* GPT technology, which counts on more training and is constantly updating, thus ensuring the use of the latest and most powerful version, less prone to hallucination. Indeed, the study started with ChatGPT powered by GPT-3.5 (which evidenced frequent hallucinations when performing mathematical operations and proved a lack of chemical reasoning), and ended up employing GPT-4 model, released in March 2023. The latest version displayed less mathematical hallucinations (within

 doi:10.20944/preprints202409.1323.v1

the frame of a K-12 chemistry and physics field), and many other advantages that will be further described (including a more trained chemical reasoning to solve some experimental problems). Other large language models such as Bard (Google), LlaMa-2 (Meta) or AWS services (Amazon), were also released, but their capacities were not comparable to those of ChatGPT at the moment [71]. Finally, the last GPT model (GPT-5) is expected to be released soon, and it has supposedly been announced to reach the Artificial General Intelligence (AGI), an AI able to pass the Turing test [87], that is an AI so developed that might be indistinguishable from a human intelligence.

2.1. Assessment of ChatGPT's Performance in the Field of Chemistry and Physics for K-12 Students

1. A set of 52 theoretical questions and problems were carefully selected to systematically ascertain the real competence of ChatGPT in the field of interest, covering the main knowledge and problem-solving skills to be acquired by 15-16-years-old students during a complete academic course. Gathering both theoretical questions and problems allowed to analyze not only ChatGPT's current strengths (textual output) but also its potential weaknesses, exploring its capacity to deal with problem-solving (combining text recognition with mathematical calculation) and also verifying the capacity to deal with inputs and outputs other than text (i.e. requesting to draw the Lewis structures of some molecules, as this is a fundamental part of the knowledge to be reached by chemistry students). The whole set of questions is available within the Supporting Information. The aim of this part of the study is not verifying if ChatGPT fails, as we already know it, but to systematically assess the amount of mistakes displayed within a real physics and chemistry test summarizing the knowledge and problem-solving skills required for a whole course, and grade it in accordance to a human scale, verifying if ChatGPT might be a trustworthy tool in K-12 science education. Finally, other parameters concerning the quality of the answer will also be taken into consideration (clarity, insight, systematicity, simplicity etc.).

2.2. Assessment of ChatGPT's Impact on Real K-12 (15-16-Years-Old) Students Learning Chemistry

In order to evaluate the impact of ChatGPT on K-12 students learning chemistry and physics without their teachers, students in the experimental group were requested to employ this tool within 4 sessions in which they had to correct their specifically designed homework and also solve theoretical or problem-solving doubts, after a previous demonstration performed by the teacher, at class. Then, two key performance indicators (KPIs) were proposed to monitor the influence of the use of ChatGPT over the students' proficiency in chemistry and physics: the users' perception on the AI as an educational tool (which was assessed by a set of questions formulated to students after each session, applying the Likert scale to address and scale the answers in the survey [88]), and student's grades (in comparison with previous term grades). Specifically, a typical five-level Likert scale design was used to measure variations in agreement, whose levels accounted for:

1. Strongly disagree.
2. Disagree.
3. Neither agree nor disagree.
4. Agree.
5. Strongly agree.

The evaluation of students' proficiency was conducted by direct comparison of students' grades through a paired sample t-test, which was performed before and after the intervention, over both the control and the experimental groups, after verifying the normal distribution of data, as previously described [89–91].

The homework was divided into four sessions focused on the "Chemical Reactions Unit", and more specifically on a topic that is usually hard to understand to the general profile of 4th ESO students, which is the fundamental chemical entity quantifying the "amount of substance", whose unit is the mole. Mole calculations at this level are related to: a) the number of particles and atoms in a specific substance using Avogadro's number, b) the macroscopic mass of the substance (including the molecular mass, or more precisely relative mass), and c) the number of moles in a gas sample

 doi:10.20944/preprints202409.1323.v1

related to the system conditions -pressure (atmospheres), temperature (Kelvin), and volume (litres)-. Within this frame, two profiles of sessions were designed:

a) *Chemical calculations (gas laws).* Sheets 1 and 2 present a concretion of the gas equation of state, in the final form of ideal gas law, related to the mole content of the gas sample. The exercises of sheets 1 and 2 request the direct and single calculation of moles, volume, or pressure from the exercise statement including the complete dataset. Each sheet includes 6 exercises.

b) *Gas or volume to mole relationships, as a more advanced learning.* Sheets 3 and 4 introduce Avogadro´s Law. Each sheet introduces a set of 6 exercises considering the calculation of a single parameter (n (mole) or V (volume)) both in the reactant and/or product species of a particular chemical reaction. Pressure (in atmospheres), temperature (absolute, in Kelvin, K) and stoichiometric factors were provided in the exercise to focus the calculations and reduce complexity. They included a brief theoretical exercise requesting a particular reformulation of the Avogadro's Law (mole to volume ratio).

The complete set of exercises and questions within the different sessions is included in the Supporting Information.

3. Results

3.1. Assessment of ChatGPT's Performance in the Field of Chemistry and Physics for K-12 Students

The performance of ChatGPT in the field of physics and chemistry for 15-16-years-old students was assessed by careful evaluation of the IA's answers to the set of 52 questions previously mentioned, which can be found in the Supporting Information, over the time of study. The score for each question relied exclusively on two parameters: the accuracy of the final result as well as the procedure to reach that outcome. Each parameter contributed a half to the total score (0.5/1), being the only inputs to assess the AI performance in the field of interest. The score for those questions including several sections was the same, so each section contributed proportionally to the final score. Finally, those resources enabling a stronger and longer-term knowledge settlement in a pedagogical manner (clarity, brevity, simplicity, use of examples etc.) were positively valued beyond ChatGPT's performance, contributing to rise the chatbot's validity as an educational tool, from a pedagogical point of view.

Preliminary tests were conducted to judge the best general prompts to be used when asking ChatGPT the different questions. The language was obviously not a problem for the tool (being a large language model): the same question was posed in Spanish and English, and the only difference was the language used to answer it (Figures S7 and S8). Besides, the straight question asked to the IA ended up with a relatively concise answer, while using a more specific prompt ("Acting as a chemistry/physics teacher, please explain…") provided more detailed but still clear answers, including accurate and illustrative examples. As a consequence, the 52 questions were evaluated by using English language and the specific prompt already mentioned. The results are summarized in the Table 1 (2023) and Table 2 (2024) and will be chronologically discussed, in order to provide a clear comparison with time.

Table 1. Results obtained by ChatGPT within the 52-questions test to assess ChatGPT's performance in the field of chemistry and physics for 15-16-years-old students in 2023.

Question	Score	Question	Score	Question	Score
1	1	19	1	37	1
2	1	20	1	38	0
3	1	21	0,67	39	1
4	1	22	1	40	1
5	0	23	1	41	1
6	1	24	1	42	1
7	1	25	1	43	1

 doi:10.20944/preprints202409.1323.v1

8	1	26	1	44	1
9	0,50	27	1	45	1
10	1	28	1	46	1
11	1	29	1	47	1
12	1	30	1	48	0,67
13	1	31	1	49	1
14	1	32	1	50	0,50
15	1	33	1	51	1
16	1	34	1	52	1
17	1	35	1		
18	1	36	1	**Final Score**	**9.3/10**

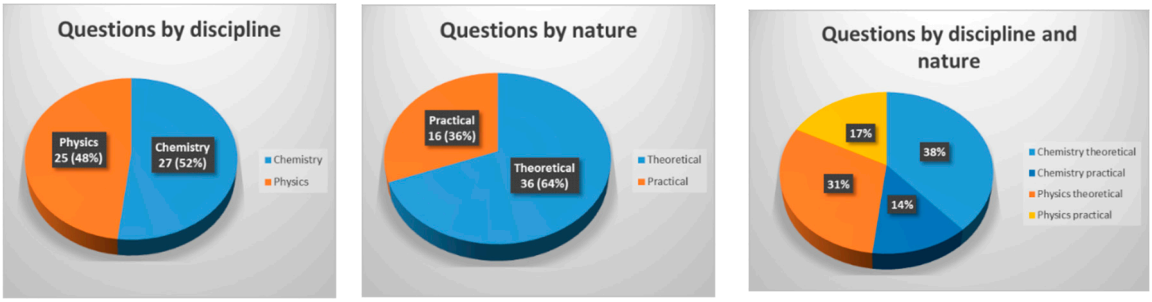
The 52 questions within the test were quite balanced according to their discipline, because 27 of them were related to chemistry, while the other 25 dealt with physics (Figure 1). According to the nature of the questions, almost 60% were theoretical queries, while 30% were experimental problems (Figure 1). Even if there was no balance, there was at least a significant number of experimental questions to show ChatGPT's competence to deal with problem-solving tasks.

While Figure 1 describes the distribution of questions by discipline and/or nature, Figure 2 displays the assessment of ChatGPT's performance in the field of K-12 chemistry and physics students, in 2023.

Table 2. Results obtained by ChatGPT within the 52-questions test to assess ChatGPT's performance in the field of chemistry and physics for 15-16-years-old students in 2024.

Question	Score	Question	Score	Question	Score
1	1	19	1	37	1
2	1	20	1	38	1
3	1	21	0,67	39	1
4	1	22	1	40	1
5	0	23	1	41	1
6	1	24	1	42	1
7	1	25	1	43	1
8	1	26	1	44	1
9	0,50	27	1	45	1
10	1	28	1	46	1
11	1	29	1	47	1
12	1	30	1	48	1
13	1	31	1	49	1
14	1	32	1	50	0,50
15	1	33	1	51	1
16	1	34	1	52	1
17	1	35	1		
18	1	36	1	**Final Score**	**9.7/10**

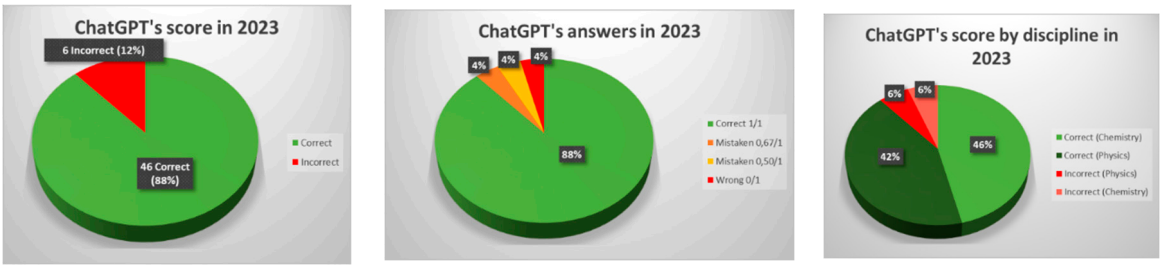
 doi:10.20944/preprints202409.1323.v1



(**a**) (**b**) (**c**)

Figure 1. Type of questions requested to ChatGPT: (a) by discipline; (b) by nature; (c) by discipline and nature.

Among the answers to those 52 questions, 46 of them were completely correct and carefully explained in 2023, that is 88% of total answers, half of them related to chemistry and the other half associated to physics. Additionally, among the 6 questions that were not correct, only 2 of them were completely wrong and scored 0 (one within chemistry and physics syllabus, respectively), while the rest were partially correct (two 0.5/1 and two 0.67/1), therefore increasing the final score from 8.8/10 to 9.3/10. Thus, ChatGPT obtained a final grade of A, which demonstrates the AI displays a quite reliable performance within the 15-16-years-old chemistry and physics syllabus, independently of the questions' nature (theoretical or problem-solving queries).



(**a**) (**b**) (**c**)

Figure 2. Assessment of ChatGPT's performance in the field of chemistry and physics for 15-16-years-old students in 2023: (a) Final score only including totally correct answers; (b) Final score including partially correct answers; (c) Final score including partially correct answers by discipline.

Some positive remarks that might be extracted from the analysis of the consistent results obtained in 2023, were:

- ChatGPT, being a language model, handled perfectly with understanding questions and providing answers in different languages (Figures S7 and S8).
- The AI, being a language model, elaborated the answers according to the literal request of information.
- The chatbot provided an answer in real-time, but it was written word by word, probably in an attempt to resemble more human, which contributed to a closer and meaningful experience for the user.
- ChatGPT was responsive to different prompts (write, act, create, list, translate, summarize, define, analyze…). In this case, the prompt "act" was exploited to request the IA to behave as a science teacher able to explain in detail the solutions to the different questions (Figure S4).

 doi:10.20944/preprints202409.1323.v1

- The chatbot took into account the context of the conversation, which might improve its comprehension of the subject being debated, and allowed to make reference to a concept or idea previously discussed (Figure S5).
- ChatGPT furnished information that was sensitive to operators as "TRUE, FALSE" (question 8, Figures S22 and S23).
- The IA could handle with not only theoretical doubts, but also with problem-solving tasks. In the latter case, the chatbot perfectly recognized and applied the values, unities and what is more important, what was being requested within the question.
- ChatGPT could return answers to several questions formulated at the same time. However, it usually provided more detailed answers when questions were divided.

According to these results and the answers included within the Supporting Information, ChatGPT displayed not only a great ability to provide correct answers to a considerable number of theoretical questions and applied problems (ensuring a remarkable performance of 9.3/10, 93%), but also clear and detailed close to human-like explanations to theoretical queries and problem-solving duties that might help students to better understand the two disciplines of study. This might imply the AI could exhibit a great competence to guide real students to a better knowledge settlement, by correcting their homework and solving their particular doubts or mistakes in real time through a positive, human-like and meaningful interaction, within an immersive and safe environment (far from judgement from teachers or peers [92]), also promoting students' confidence and self-regulation.

Considering now the 6 incorrect answers in 2023, they were balanced by discipline, being 3 of them related to chemistry and the other 3 related to physics. However, there was no balance by nature, as there were 2 issues with practical problems and only 1 issue concerning theoretical questions within chemistry, while the opposite situation happened with physics wrong answers. Anyhow, no tendencies concerning the theoretical or practical nature of the incorrect answers could be extracted.

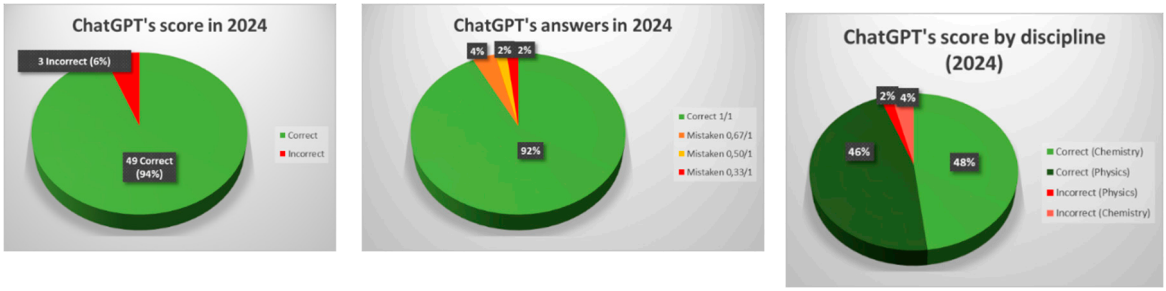
Furthermore, the main problems encountered by ChatGPT in 2023 focused on its own inability to recognize or produce images at the moment (question 38, Figure S64, and questions 9 and 21, Figures S24, S42, S43, respectively), even if an accurate textual description was provided instead, and only on one occasion it found difficulties to solve simple mathematical calculations (question 50, Figures S81 and S82), even if the theoretical procedure and the substitution of numerical values in the equations were correct. Finally, the AI found also some troubles when predicting periodic properties of elements (the direct consequence of electronic configurations) to order some elements according to certain properties, such as radius and reactivity (question 5, Figure S17), as well as discussing about the type of energy (kinetic or potential) exploited in several sources of energy, specifically in tidal energy (which might exploit both kinds of energy, even if the chatbot was forced to decide one of them). In summary, the AI problem to process images as inputs or outputs involved 3 among the 6 incorrect questions in 2023. While this issue might only find a solution through an application redesign, the rest of the troubles might be solved with better training of the GPT model, allowing improved and more accurate answers.

A brief summary of the chatbot weaknesses in 2023 that might be extracted from the analysis of the mistaken answers were:

- Being a language model, ChatGPT could not recognize images as proper inputs. When the query could not furnish all the inputs required to understand the question, the user was forced to develop an alternative code to introduce the lacking data (such as that in question 1, involving a customized notation created in real-time to make the AI understand how to recognize the atomic mass and atomic number of some isotopes that were provided within the question, Figure S5).
- The chatbot could not create image as outputs even with GPT-4 (i.e. question 21), though the textual description that was offered instead was very clear, illustrative and correct.
- The AI, being a language model, encountered some problems with mathematical calculations (question 50, Figures S81 and S82). Even if they were more frequent in GPT-3.5 model, occasional mathematical hallucinations persisted in GPT-4 model.
- The chatbot did not handle correctly all the periodic properties of elements.

 doi:10.20944/preprints202409.1323.v1

Despite these results, ChatGPT was further evaluated in 2024, and surprisingly the chatbot was even more competent than before. Among the 52 questions, 49 of them (94%) were completely correct (Figure 3), and there was no completely wrong answer (0).



(**a**) (**b**) (**c**)

Figure 3. Assessment of ChatGPT's performance in the field of chemistry and physics for 15-16-years-old students in 2024: (a) Final score only including totally correct answers; (b) Final score including partially correct answers; (c) Final score including partially correct answers by discipline.

The 3 questions that were partially correct increased the final grade from 94 to 97%, still scoring A but displaying a significant performance improvement in 2024. All the theoretical questions were correctly answered by ChatGPT in 2024, including that regarding the properties of elements according to their position within the periodic table (question 5, Figures S17-S17b), that concerning the kinetic and also potential energy of tidal energy (question 48, Figures S77-S79b) and even the problem of handling images as inputs was solved (question 38, Figure S64-S65b), recognizing the vectorial character of Force within an image provided by the user.

Despite the verified improvement with training, the AI still exhibited some difficulties with 3 practical questions, specifically with handling images as outputs (questions 9 and 21, Figures S24, S42, S43, respectively) and also with some mathematical calculations (question 50, Figure S81-S82b). Some Lewis structures were clearly improved, and the textual description was perfect, but the final image was still confusing (Figure S24b). The same stood for the energy diagram requested in question 21 (Figure S43b): even if the scheme indicated in parentheses that the energy of reaction products was lower, the drawing placed the energy of reactants in a higher position than that of reaction products, which could be confusing for students. Besides, reactants and products were in the same level within the x axis (reaction progress), therefore the student might not appreciate the variation of energy during the reaction progress in a clear way, which is the aim of that part of the question.

Regardless of these minor problems (many of which were duly addressed by a better training of ChatGPT in 2024), the consistent, exhaustive, and positive results obtained within the test demonstrated the remarkable performance of ChatGPT to answer both theoretical and problem-solving questions (9.7/10, 97%) , scoring A, thus being trustworthy for K-12 physics and chemistry students. Furthermore, the first part of the study unveiled the positive resources (beyond performance) to enhance students learning process: clarity and a high level of detail and organization in the real-time answers provided in a time-independent manner, as well as a human-like meaningful interaction with students, which provided them with complete freedom to exploit this tool to find real-time answers to their particular doubts when they are studying or doing homework, supporting them in a way no other educational tool would allow under these circumstances (when teachers are unavailable). These advantages pave the way for a potential use of ChatGPT assisting teachers in their task of mentoring real students.

3.2. Assessment of ChatGPT's Impact on Real K-12 (15-16-Years-Old) Students Learning Chemistry

The impact of using ChatGPT as a virtual mentor on real 15-16-years-old students learning chemistry when teachers are unavailable has been assessed through an empirical interventional study performed in a real school, monitoring two KPIs: the users' perception on the use of the AI as an educational tool (evaluated by a set of questions formulated to students after each session, see

 doi:10.20944/preprints202409.1323.v1

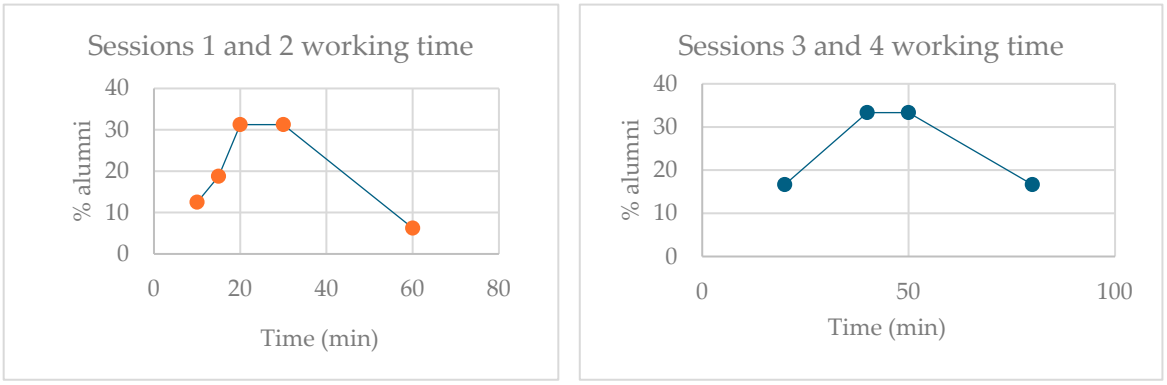
Supporting Information, applying a typical five-level Likert scale), and students' proficiency (by comparing students' grades before and after the intervention, first and second term, respectively).

The study comprises the analysis of several exercise sheets including the main calculations concerning chemical reactions, as presented in the former section, and a subsequent use of ChatGPT to verify the correction of the homework, as well as solving any mistake or doubt. During the class, before the release of the homework sheets, the teacher corrected at least one problem with the help of ChatGPT, thus the students had an initial guide to the use of this tool with autonomy (prompts, possible mathematical hallucinations etc., trying to promote students' critical thinking). The AI followed a general procedure to solve problems that basically consisted of 1) identifying the data (including unities) and the unknown factor among pressure (P), volume (V), amount of matter (n), and temperature (T), 2) determining the formula required to solve the problem, and 3) performing the substitution of real values within the formula (sheets 1, 2) or basically identifying species in the chemical reaction, associating data to them, and performing the substitution of real values within the formula (sheets 3, 4). Yet in the preparatory class, the use of ChatGPT 3.5 released some minimal and basic calculation errors that could be solved with the human factor (help of the teacher and careful surveillance of students), boosting students' critical thinking but conditioning to a certain extent their initial opinion on the tool. The text of the solution improved slightly from trial to trial, and this was the furthest scope of ChatGPT 3.5's approach. Upon completion of the study, some questions were tested again with ChatGPT 4.0, reducing the issues with mathematical calculations and improving the chemical ability to solve problems and teach students, as the AI was now able to establish relations among the chemical species in the chemical reaction (i.e. assignment of a given formula to a reactant, or a product), and determine not only their stoichiometric coefficients, but also other important conditions such as limiting and excess reactants. In fact, this means a huge step forward in solving chemical reactions exercises in comparison with the previous version, ensuring a stable and sure pathway to the correct solution.

Once the four sessions data sheets were collected, the student's tasks were corrected, evaluated, and qualified. The student's performance questionaries were registered to be evaluated after the sessions' completion.

3.2.1. How Long Did It Take to You to Complete the Session?

Considering question 1, Figure 4 represents the time spent by students to solve the homework (without the AI), displaying a tendency with a Gaussian shape, centered in 24 minutes for sheets 1, 2, and 47 minutes for sheets 3, 4, which is reasonable as the latter entailed a much higher calculation load.



(**a**) (**b**)

Figure 4. Question 1 concerning the working time devoted to complete sessions (a) 1 and 2; (b) 3 and 4.

 doi:10.20944/preprints202409.1323.v1

3.2.2. In What Aspects of the Session Have You Found More Difficulties?

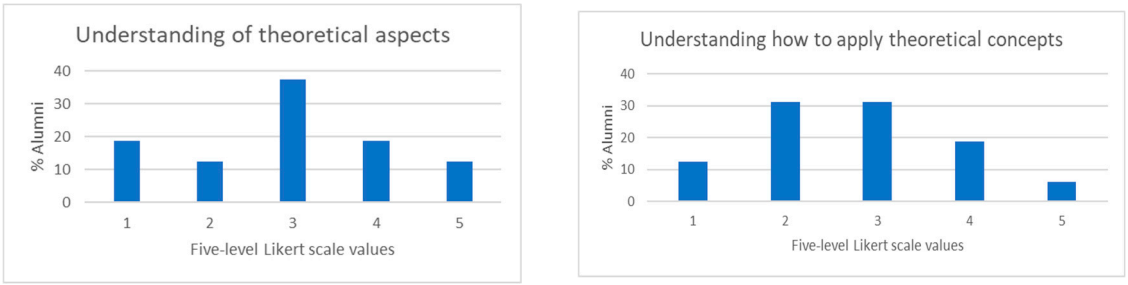
The most common response to this question was the effort to understand chemical concepts and consequently, the way to apply them in a real problem.

3.2.3. Rate Your Level of Agreement (1: Strongly Disagree, 2: Disagree, 3: Neither Agree Nor Disagree, 4: Agree, 5. Strongly Agree) with the Following Statements:

3.2.3.1. You Have Understood the Theoretical Concepts

3.2.3.2. You Know How to Apply the Theoretical Concepts

As expected, the results to question 3 (Figure 5) supported the most common answer of students to question 2 ("In what aspects of the session have you found more difficulties?"), proving that the majority of students (69%) did not agree with having understood theoretical concepts, and then being able to apply them within a problem-solving task implying some calculations (75%), without the teacher's or the AI's help. Because of this, the use of ChatGPT to solve doubts and to correct the homework assignments arose as a potential solution for those students facing problems to complete the sheets exercises, improving students' understanding of theoretical concepts in real time, and guiding them to apply those concepts within real problems, in a very clear and detailed manner.



(**a**) (**b**)

Figure 5. Question 3, regarding students' perception of their understanding of: (a) theoretical concepts; (b) the application of theoretical concepts.

Finally, question 4 was only requested to students having employed ChatGPT.

3.2.4. Rate Your Level of Agreement (1-5) with the Following Sentences:

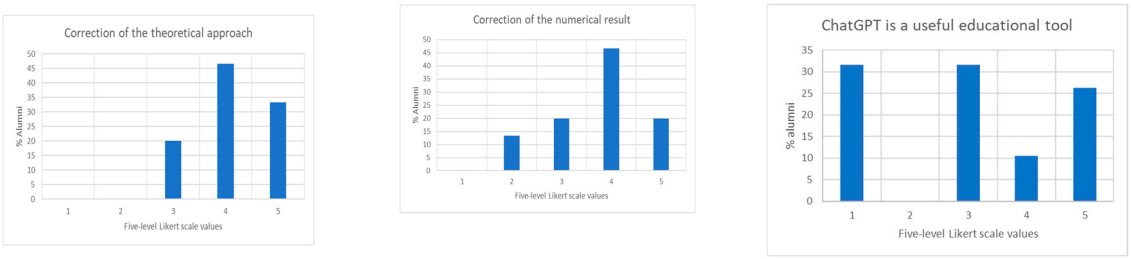
3.2.4.1. The Approach Offered by ChatGPT to Solve the Exercise Is Correct

3.2.4.2. The Numerical Result of the Exercise Provided by ChatGPT Is Correct

3.2.4.3. ChatGPT is Useful as a Complementary Educational Tool (For Solving Theoretical Doubts or Correcting Problems) in the Absence of a Teacher

Students having used the chatbot then evaluated the AI ability to solve chemical problems. As expected, Figure 6a points out the positive perception of students about the competence of ChatGPT (even if GPT-3.5 model was firstly used) to define the theoretical approach to solve the exercise, displaying a distribution of values centered at 4 (agree).

 doi:10.20944/preprints202409.1323.v1



(**a**) (**b**) (**c**)

Figure 6. Question 4, concerning the correction of (a) the approach to solve the exercise; (b) the numerical result; and (c) the usefulness of ChatGPT as an educational tool.

In a similar manner, the results for question 4.2 (Figure 6b) show a high degree of acceptance of the statement, that is, students appraised the AI capacity to provide correct numerical results most of the time, even if there were some mathematical hallucinations, more frequent with GPT-3.5 model (and still present to a lower extent in GPT-4 model). Those errors of GPT-3.5 model might be the reason for a wider distribution in Figure 6b, tending towards lower values because some students (13%) disagreed considering ChatGPT offered correct numerical results (in general), being the average still centered at 4 (agree). The mathematical errors might easily be detected by a human user, and this fact could be positively exploited by both students and teachers. Concerning students, the potential presence of hallucinations might improve their attention as well as their critical-thinking ability. Regarding teachers, the risk of finding a mathematical mistake prevents students from negative behaviors like the direct copy of results, promoting a good use of the AI based on following ChatGPT's guidance to the solution through a perfectly detailed theoretical approach. In conclusion, ChatGPT might be perceived as a patient and wise mentor correcting homework and solving students' particular doubts in real time, displaying occasional mathematical distractions that can easily be detected by students, all of which provides a meaningful and positive interaction with K-12 students, improving their learning process.

The questionary concludes with question 4.3 (Figure 6c), which provides an overall impression of students' perception about the AI as an educational tool. Only a 37% of students appreciated ChatGPT's ability to guide them towards the right pathway to solve/correct a problem (agreeing (4) and strongly agreeing (5) with the utility of the AI as an educational tool). Thus, there is a considerable number of students which were not sure about ChatGPT's usefulness (31%) or strongly disagreed with its utility (32%), which was comprehensible considering they started the study testing GPT-3.5 model. These students probably put more focus in the problem of mathematical hallucinations rather than valuing the AI capacity to detail a perfect theoretical approach to solve any chemical exercise. Furthermore, there were additional issues underneath that might be conditioning this result, which were: 1) the limited time of use and the sooner evaluation of the AI (right after the completion of each homework session), and 2) the lack of objective indicators for students to measure the improvement in their learning process (such as the increase of their proficiency in chemistry). Thus, the same survey was repeated after a whole term of evaluation, in order to notice any remarkable difference in students' perception on AI. Besides, this KPI was not the only one employed to monitor ChatGPT's impact on 15-16-years-old students learning chemistry.

Students' proficiency before and after the intervention was also assessed by direct comparison of students' grades after one term of AI use with those obtained the previous term (with no AI assistance), through a paired sample t-test, which was performed before and after the intervention, over both the control and the experimental groups, after verifying the normal distribution of data. The results of the data analysis are summarized in Table 3 and Figure 7.

The general overall marks of these students have improved since the utilization of the AI tool, although grades are indeed a really complex factor. Within the control group, even if the average marks slightly improved on the second term (one point out of ten, Figure 6), the data analysis revealed that there were no statistically significant differences between the students' grades before and after the intervention, with 95% confidence (which was set for the study), as the p-value

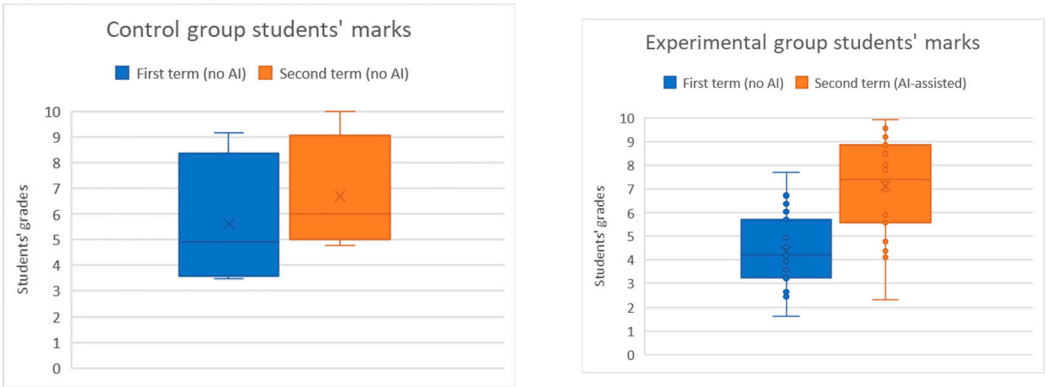
 doi:10.20944/preprints202409.1323.v1

associated with the contrast statistic of paired samples Student's t test (p = 0.29) was greater than 0.05 (5%). This might be expected, as the control group did not employ ChatGPT as a virtual mentor, so these students could not exploit all the advantages of the AI use proposed within the study.

On the contrary, the experimental group displayed statistically significant differences between the students' grades before and after the intervention, with 95% confidence, as the p-value associated with the contrast statistic of paired samples Student's t test (p = 1,67.10-6) was smaller than 0.05 (5%). The students' mean scores in the experimental group improved almost three points out of ten (Figure 7) after one complete term using ChatGPT as a virtual mentor. Besides being statistically significant, this remarkable improvement in students' proficiency might be due in a large extent to the use of ChatGPT as a virtual mentor in the absence of their teacher, because the improvement 1) was almost three-times higher than that of the control group, and 2) was manifested by 90% of students in the experimental group, independently of their level of proficiency.

Table 3. Assessment of students' proficiency by comparison of students' grades before and after the use of ChatGPT as a virtual mentor during one term, through paired sample t-tests applied over the control and the experimental groups.

Control Group	Before	After	Experimental Group	Before	After
Mean	5,62	6,69	Mean	4,37	7,11
Standard deviation	6,8225	5,2588	Standard deviation	2,5190	4,3867
Observations	4	4	Observations	19	19
Pearson correlation coefficient	0,7697		Pearson correlation coefficient	0,5951	
Hypothetical difference of means	0		Hypothetical difference of means	0	
Degrees of freedom	3		Degrees of freedom	18	
t statistic	-1,2654		t statistic	-6,9602	
P(T<=t) one-tailed test	0,1475		P(T<=t) one-tailed test	8,3829E-07	
t critical value (one-tailed test)	2,3533		t critical value (one-tailed test)	1,7341	
P(T<=t) two-tailed test	0,2951		P(T<=t) two-tailed test	1,6766E-06	
t critical value (two-tailed test)	3,1824		t critical value (two-tailed test)	2,10092204	



(**a**) (**b**)

Figure 7. Second KPI: students' grades in the first and second term (before and after the intervention) for: (a) the control group and (b) the experimental group.

The students with high level of proficiency improved their marks significantly, but the main differences were observed for students with medium and low level of proficiency, as their grades improved remarkably. Students with a failed first term (marks between 3 and 5) got second term grades between 6 and 9.55, and the most shocking example, the student with the lowest mark in the first term (1.61, showing great difficulties to follow the class level), improved her knowledge during

 doi:10.20944/preprints202409.1323.v1

the second term using the AI to finally get a 4.78 (slightly higher final grade, 5.52, with additional contributions from other tasks of no interest for this study), reaching the minimal acceptable level of knowledge for a student in Spain (5/10). These particular cases reflected that the average level of proficiency of the class climbed from low-medium to medium-high, with the only aim of this educational tool assisting the teacher.

Finally, before the end of the final term, the students' perception on ChatGPT's utility as an educational tool was again assessed, in order to compare with the previous results (Figure 8). At that moment, the students had used the AI much longer time (two terms), and they also counted on several objective criteria to assess whether their proficiency in chemistry had increased or not (their own knowledge, which could still be somehow subjective, and their grades, which were objectively assigned by the teacher).

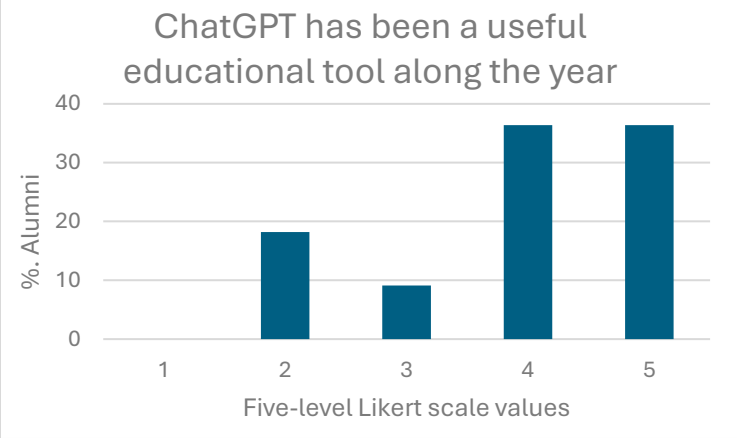


Figure 8. Question 4, concerning the usefulness of ChatGPT as an educational tool, after two terms using the AI.

As expected, the overall perception of students about the use of ChatGPT as an educational tool after two terms of use was significantly more positive than previous results: 70% of students agreed or strongly agreed considering the AI as a useful educational tool, instead of the initial 37%, that is, almost the double of students providing a positive perception of ChatGPT. According to what was commented before, a higher time of use and objective indicators of students' proficiency improvement (their real proficiency, and their grades) might account for this change in students' perception of the AI, realizing that ChatGPT had been an efficient educational tool to boost their proficiency in chemistry in a short time.

Beyond the data analysis, the survey allowed to gather some students' opinion that went beyond a positive perception: they agreed the homework sheets and the work with ChatGPT had been crucial for the improvement of their learning process. When students faced the homework assignments, teachers were unavailable, and regretfully, many of them could not count on parents or any other tutor able to help them with their doubts in real time (which is by the way a common problem for teen students, which are allegedly supposed to be more independent). Those students found themselves suffering a lack of surveillance, advice and/or support from a meaningful human being around them that no one could fill in. However, that feeling disappeared after one term of using the AI at home, because students now could count on ChatGPT to provide them with the real time support (both technical and emotional support) they needed. Thus, the study empirically demonstrated the remarkable abilities of ChatGPT to mentor 15-16-years-old students through a human-like, meaningful and personalized interaction, which not only ensured a remarkable improvement of students' proficiency, but also contributed to promoting a feeling of care and support some students lacked[64].

 doi:10.20944/preprints202409.1323.v1

4. Discussion

The results obtained within the present exploratory study allowed to address the original research questions in the frame of K-12 science education:

RQ1: Does ChatGPT provide a trustworthy time-independent learning experience to K-12 students, when teachers are unavailable?

Previous studies of ChatGPT performance unveiled quite different results, even within the same field of application and target population [58]. For instance, ChatGPT returned neither reliable nor accurate information concerning anatomy for university studies [60], it displayed poor results (45% of correct answers) for MSE (the main specialization exam for medicine in Türkiye) [93], and finally it also provided a high level of concordance, accuracy, and insight on the United States Medical Licensing Exam (USMLE) [48]. Several factors might account for this variability across the literature, besides the knowledge level and the AI model (which are the same in these cases): systematicity, type of questions and the level of encoding. The first study attempted a quick assessment of the AI performance within the field of anatomy by means of a few questions, while the rest of them were more exhaustive and systematic, i.e. evaluating the AI performance through three tests of around 300 multiple-choice questions. The second study provided the AI with multiple-choice questions, while the third study used not only multiple-choice but also open-ended questions. Besides, the latter study put special attention to encoding, including both multiple-choice single answer with and without forced justification prompting. Surprisingly, the most exhaustive and systematic study, which included several types of questions beyond multiple-choice ones and paid particular attention to encoding, concluded the AI performed with a high level of concordance, accuracy, and insight on the United States Medical Licensing Exam. This suggests that any assessment of current AI capabilities within certain field of knowledge should take (at least) all those variables into account, that is, knowledge level, AI model, systematicity, types of questions (not only multiple-choice but also open-ended questions) and encoding.

In the field of science education there are some publications exploring the ability of the generative artificial intelligence to answer a few theoretical questions [68,70,71], but more systematic studies paying attention to the previously described variables would increase the knowledge in the field.

In this study, the lower level of knowledge (restricted to K-12 science education) and the higher AI model planned to be used during the experiment (GPT-3.5 at first, and then GPT-4) made the authors foresee a positive performance. Furthermore, a special attention was dedicated to performing an exhaustive and systematic assessment, including more than 50 open-ended questions of K-12 level chemistry and physics, and an adequate encoding. No demographic, cultural, linguistic, temporal, or ideological/political biases were detected within the answers, probably because the questions belonged to the scientific field, leaving aside those sources of potential bias from the language model. The exhaustive and systematic study revealed that ChatGPT provided answers with remarkable accuracy and insight to theoretical questions, and a clear, detailed and well-organized procedure to find the solution to problem-solving questions. The consistent and positive results obtained through this performance test (9.3/10 in 2023 -with GPT-3.5 model-, and 9.7/10 in 2024 -with GPT-4 model-) demonstrated ChatGPT provided trustworthy answers (97%) in the field of K-12 science education, despite some limitations that have been considered and included within the test design (half of which have been duly addressed by OpenAI after a GPT model update and more model training). Thus, a realistic view of current ChatGPT performance within the field of interest has been provided to the community, which is in line with previous publications considering similar variables within different fields of education [48,72]. These results suggested ChatGPT might be a reliable tool to help K-12 students learning science to reinforce theoretical knowledge and enhance their problem-solving skills with no time and location restrictions, when teachers are unavailable.

RQ2: Can ChatGPT create meaningful interactions with K-12 students?

Even if there is still a limited understanding of what meaningful interactions are, from a holistic perspective, some authors recently summarized the main factors contributing to meaningfulness

 doi:10.20944/preprints202409.1323.v1

across different cultures, taking into account today's media landscape [94]. Those aspects included the partner and what happens before, during, and after the interaction [95], the number of interactors [96], the activity during the interplay [97], as well as the communication medium [98], synchronicity [99,100] and motivation/engagement. According to this study, factors regarding the interaction characteristics had more influence over the meaningfulness than the communication channel [95]. Besides, among the different communication media, text and instant messaging as well social media/network had higher influence than calls, video calls and even face-to-face interactions (which displayed the lowest coefficient from linear regression analysis).

If the interaction of users with ChatGPT could be examined as an analog of the interaction of two human beings, from a theoretical perspective, ChatGPT might be taking benefit from some of those factors promoting meaningful interactions with users, to a certain extent. First, the interplay involves only one interactor (and some authors suggest people might find more meaning in small groups [97]). Besides, the communication medium is an instantaneous textual message conversation, which is the best option among the different media, and one of the activities providing maximal meaning to the interaction [95]. Furthermore, the conversation aim is studying, which is also another activity driving to an interaction of maximal meaning. Finally, synchronicity is another strength of ChatGPT, which might enhance the meaningfulness of the interaction through "amplification effects", related to more vivid memories, motivation and engagement [95,100,101]. Thus, the AI puts special effort into interacting with the user like a human being would do, not only displaying answers in real time, but more precisely typing them letter by letter, word by word, sentence by sentence, contributing to a closer and more realistic interaction. In conclusion, ChatGPT counts on the required resources to be able to create meaningful interactions with the user, from a theoretical point of view.

During the experiments to assess ChatGPT's performance in the field of chemistry and physics for K-12 students, and those aimed at evaluating ChatGPT's impact on real K-12 (15-16-years-old) students learning chemistry when their teachers were unavailable, both the authors and the students verified the AI remarkable ability to explain scientific theoretical concepts and guide students to solving problems like a real human would do, following conventional well-organized procedures in a clear way, and providing real-life examples promoting a deeper and longer-term understanding of the subject. Both authors and students corroborated the interaction with ChatGPT resembled quite similar to interacting with a real person with a deep knowledge in the field (K-12 physics and chemistry). Many aspects of ChatGPT were responsible for this resemblance, such as the real-time instant-messaging-like communication medium (contributing to both synchronicity, motivation and engagement,) with only one interactor, typing the answer as a human-like would do, the ability of the AI to take the context and previous discussion into account, very open and versatile prompts enabling the user to request ChatGPT to answer in a particular framework (i.e. acting as a secondary school teacher), and its ability to provide rich answers adapted to those different requested behaviors. All these experimental aspects of ChatGPT allowed creating meaningful interactions with the user.

Besides the functional advantages of the AI enabling a meaningful interaction with students, a different facet enriching the meaningfulness of the interaction might also be considered, such as the emotional support the AI might provide to students. Several authors recently claimed that ChatGPT cannot offer emotional support nor enhance critical thinking and problem-solving skills in science learning [68], despite its impressive capabilities (assessing, grading, guiding and recommending students), but the discussion kept within the theoretical plane, mainly referring to the fact that ChatGPT is not able to substitute the role of teachers (which is true).

However, the present study proposed the AI might assist the teachers in their role of mentoring students (not substituting them), and thus has experimentally verified that ChatGPT displayed not only functional but also social features. In fact, the AI was able to satisfy individual needs in real time (K-12 scientific doubts), within a safe environment (far from peers, parents' or teachers' judgement [93]) and without time or location restrictions, providing both fast answers (of mainly any domain) to users' particular questions as well as the interactive and personalized support desirable from the perfect assistant [64]. Thus, the AI-human interaction promoted users' autonomy and productivity, which in agreement with Maslow [101] and the Relationships Motivation Theory [102] contributed

 doi:10.20944/preprints202409.1323.v1

to satisfy their physiological, emotional, security, dignity and self-actualization requirements, even promoting a feeling of care, support and social camaraderie [64]. Emotional support is usually provided by family, significant others, friends, colleagues, counselors, therapists, clergy and support groups, but also by online groups or even social networks. In this case, the meaningful interaction established between the AI and the human user enabled ChatGPT to provide students with a kind of emotional support that, according to bibliography [103], might bring students reassurance, acceptance, encouragement, and sense of being cared.

This was not only verified by students' positive perception on the AI after two terms of evaluation, but also through some students' opinion that was obtained within the survey, corresponding to students counting on no parent nor tutor able to help them with their doubts during their time to solve the homework assignments. They highlighted how important was counting on the AI to solve their doubts in real time for their learning process. The meaningful interaction established between ChatGPT and those students in a safe environment, in combination with the remarkable technical support provided by the AI in a time- and location-independent manner, promoted student's sense of care and reassurance over time (two terms), which made them finally recognize the emotional support they felt knowing they could count on ChatGPT to help them.

In conclusion, ChatGPT was able to create meaningful interactions with K-12 students (15-16 years old), and the emotional support provided by this singular human-AI interaction reinforced its ability to assist the teachers when they were unavailable within the virtual mentor role proposed in the study.

RQ3: What is the real impact of using ChatGPT as a virtual mentor on K-12 students learning science when teachers are unavailable?

The objective of using ChatGPT on K-12 students learning science when teachers are unavailable is to help students correcting homework assignments, solving doubts and guiding them towards a better understanding of the lesson and a stronger and longer-term settlement of knowledge (technical support), while improving students' sense of care and reassurance (emotional support). Thus, the impact of this approach might be verified by a raise of students' knowledge and skills, and the evolution of students' perception of ChatGPT as a useful educational tool.

The increase in students' knowledge and skills was experimentally assessed by a quasi-experimental analysis [104], an empirical interventional study avoiding randomization able to determine the causal effects of an intervention on the target population. Even if an experimental analysis would have provided stronger causal effects, randomization in a small group of students (a class) might have created groups with unbalanced level of proficiency, therefore it was avoided. Future studies will soon be carried out with a higher number of students, considering several classes and schools from different regions/countries, and thus an experimental analysis will be chosen to develop a large-scale assessment.

Once the remarkable performance of the chatbot in K-12 chemistry and physics was demonstrated, and the meaningful human-AI interaction could be verified, the effectiveness of the proposed educational approach was monitored through two different outcomes: the evolution of students' grades before and after the intervention and students' perception on the AI as an educational tool.

According to bibliography, grades offer a limited capacity to evaluate the students' level of proficiency (due to a generalized lack of standardization across institutions or even nations). Despite this, grades are still one of the most frequently used indicators, trying to systematically take into account both theoretical knowledge and applied skills and competencies through achievement tests [105]. Besides, recent studies demonstrated that high-school grades might be a stronger predictor of college grades than standardized tests (because they are thought to capture both students' academic and noncognitive factors that play a role in academic success, such as perseverance and a positive mindset) [106]. In conclusion, grades might not only be considered as reliable indicators, but also good predictors of future performance.

On one hand, the present study demonstrated that students' grades average in the experimental group improved 30% after one term using ChatGPT to correct their homework assignments and also

 doi:10.20944/preprints202409.1323.v1

to solve theoretical and problem-solving doubts (with respect to their grades in the previous term, with no AI help). Besides being statistically significant, the improvement in students' proficiency was almost three-times higher than that of the control group and was manifested by 90% of students in the experimental group, independently of their level of proficiency. Considering students' perspective within the experimental group, the students with lower level of proficiency (displaying great difficulties to follow the lessons) were able to pass their exams (some of them reaching good marks), while the students with higher level of proficiency still increased their grades. To sum up, the class grades' average improved from low-medium to medium-high level of proficiency, with the only additional help of ChatGPT as an educational tool assisting students when teachers were unavailable.

On the other hand, students' perception on the AI varied over the course of the study. First reason is they handled evolving versions of ChatGPT: they started using GPT-3.5, displaying more limitations and hallucinations, and they ended employing GPT-4, which addressed most of those problems. Furthermore, time allowed students to provide a more meditated perception on the AI. After one term using ChatGPT, students counted on not only perspective, but also quantitative inputs to verify if the AI had been a useful educational tool for them, or not: their own proficiency, and grades. Both reasons made the students' perception on the AI as an educational tool evolve from an average of 3 (neither agree nor disagree), with a third of students strongly disagreeing (1), towards an average of 4.05 (agree), with most students agreeing or strongly agreeing (70%).

Both indicators clearly demonstrated in this quasi-experimental analysis that the intervention was successful: ChatGPT was capable to provide students with the technical and emotional support they required when teachers were unavailable, so that their grades and perception on the AI usefulness as an educational tool increased significantly after only one term.

Similar approaches have already been proposed in the field of education [56,73,107], but no experimental analysis was performed, and the lack of empirical results was highlighted. To the best of our knowledge, this is the first experimental assessment of the impact of using ChatGPT on real science students' learning outcomes, monitoring academic achievement and perception on AI as an educational tool assisting teachers in their task of mentoring students when they are unavailable (i.e. at home). This finding supports previous conclusions of AI chatbots enhancing learning performance, verified in other educational areas such as language learning [108–110] and even through a meta-analysis [61], claiming a positive impact of AI chatbots in several learning outcomes.

Previous publications [108] described potential limitations of ChatGPT, including: 1) an effectiveness that has not been fully tested, 2) the quality of data used to train the AI, and 3) the complexity of the tasks to be performed by ChatGPT. The present study demonstrated ChatGPT might overcome those limitations in certain context. Restricting the knowledge level to well-established science, such as the chemistry and physics for K-12 students, the difficulty of the tasks to be performed by the AI was moderate to low, and thus 2) the quality of the well-known data used to train the AI in this context was noticeable (as the few mistakes of AI during the performance assessment were related to a mathematical hallucination and the handling of image inputs/outputs, all of which should be improved and was not related to the quality of data). Those reasons and the powerful AI model employed (GPT-4 in the end) might explain the remarkable theoretical performance obtained (97%), within the specified context. Once ChatGPT was applied to mentor students while teachers were unavailable, such an educational tool promoted a 30% increase in the grades of students within the experimental group. Surprisingly, this finding is not aligned with the conclusions of previous studies assessing the effect of AI chatbots across different educational levels. According to Garzón and Acevedo [111], the support from AI chatbots significantly improved University students' learning outcomes, but there was no significant effect on primary school and secondary school students' outcomes. Some studies [112] verified that the lower language competency of primary school students might hamper an effective interaction with AI chatbots. However, the present study demonstrated not only that secondary school students are able to create a meaningful and effective interaction with ChatGPT (15-16 years old students count on high enough language competencies), but also that the impact of the AI on secondary school students was

 doi:10.20944/preprints202409.1323.v1

remarkable. Maybe the improvement in both the AI training during the present study (2023 and 2024) and its ability to meaningfully interact with the user accounted for this positive change of trends.

In order to evaluate these results, it is important to describe the educational context. Spanish educative law (LOMCE at the moment, evolving towards LOMLOE) established a curricular system for K12 students, which particularly for 4th degree of compulsory secondary education (hereinafter 4th ESO), introduced basic concepts as molar concentration, ideal gas law, stoichiometry, and simple calculations regarding mass conservation law, or limiting reagent. 4th ESO is the last course of compulsory education system in Spain, and even if "physics and chemistry" subject is optative at this level, it is frequently chosen among the different possibilities by a high number of students. However, a significant part of them will not undergo to further bachelor or university studies, so the teacher must do their best effort to keep motivation at a high level. Therefore, blended approaches involving AI like the one proposed in this study might contribute to motivating and engaging students in an efficient manner. Besides this advantage, there is another important issue to consider in the temporal educational context: the academic years 2019/2020, 2020/2021 and 2021/2022 included strong educational changes due to the COVID-19 pandemic, and despite the employment of new TIC resources, most of the students still showed gaps in their STEM learning process. Thus, ChatGPT might not only raise students' motivation and engagement, but also assist teachers in their tasks of guiding and supporting students (when they were unavailable), solving their particular doubts and even bridging those educational gaps in a time-independent manner, allowing all students to catch up and gain confidence. Therefore, the obtained results demonstrated the AI might become an educational tool able to democratize a higher level of knowledge acquisition, without the need of parents/tutors/private teachers help, thus promoting students' autonomy and security.

This idea aligns with the conclusions of previous studies, emphasizing the intertwined evolution of society, education, and technology [113]. New opportunities for a more inclusive, accessible and effective education might arise when using ChatGPT as an assistive technology that automates communication in this field. While the use of AI could respond to societal needs (such as providing students with assistance in a time- and location- independent manner), ChatGPT might also have an impact on society and education, and this impact might promote the development of more responsible technological advancements, at the same time including newer learning opportunities and ever closer AI-human interaction [114].

Many advantages of the use of ChatGPT as an educational tool might explain the obtained results, among them the rise in students' confidence (solving all the doubts the student needs in real time, even from their backgrounds, to understand theoretical concepts and also problem-solving exercises, correcting their homework and localizing potential mistakes), an increase in their motivation and engagement (testing a disruptive educational tool which implies a meaningful interaction with a virtual entity, with certain degree of gamification), and the benefits from virtual mentoring, such as completely personalized learning (students ask exactly what they might need), location- and time-independent learning (students might use the AI at any place, with any media -pc, mobile phone-, and any time they want to learn), and the meaningful interaction with the chatbot (as previously discussed in RQ2), ensuring a long-term knowledge acquisition (functional support) and improving students' sense of care and reassurance (emotional support) etc. Again, these conclusions align with previous research providing reasons to explain why students assisted by AI could increase their learning performance: an increased confidence, motivation, self-efficacy etc. [61].

The main limitation of this study lied in the limited number of students participating within the interventional experiment (one class of science in a single school of Benaguasil, Valencia, Spain). However, the study aimed at providing the science education community with the first proof of the positive impact of using ChatGPT as virtual mentor of K-12 students learning science, when teachers were unavailable, so the allowance to conduct such a new interventional study was restricted to only one class. After the positive results obtained within this study, future experimental analysis (randomized studies) including a higher number of students and a broader diversity (considering age, gender, nationality, ethnicity, ability, religious, socioeconomic, experiential, sexual orientation

 doi:10.20944/preprints202409.1323.v1

or geographical diversity) will soon be conducted to gain a more complete understanding of the AI impact on education.

Finally, two potential biases might have influenced the results: the Hawthorne effect (or the modification in human behaviour when the individual being observed is aware of being studied) and the John Henry effect (the change in human behaviour in individuals belonging a control group, trying to compensate for their apparent disadvantage). In this case, no solutions to those biases are foreseen for future studies, as individuals participating within such experiments must use or avoid using AI, so they are obviously aware of being studied, and also conscious of belonging the experimental or control group.

Despite the limitation and potential biases, the study suggested ChatGPT might be a useful educational tool able to provide K-12 students learning science with the functional and emotional support they might need, democratizing a higher level of knowledge acquisition with no additional help from parents/tutors, and promoting students' autonomy, security and self-efficacy. The results probe ChatGPT's experimental capacity (and huge potential) to assist teachers in their mentoring tasks, when they are unavailable, paving the way for future studies allowing to get a more realistic perception of the AI impact on education.

As a final recommendation, students and professors should be exhaustively trained in order to unleash the vast potential foreseen for the AI in the field of education. This would allow them to better exploit the benefits from AI technologies over time, and also to gain insight into their potential risks, biases and limitations.

5. Conclusions

Current reviews on the impact of the AI on education claim their main limitation is the lack of empirical studies assessing the effect of using AI on students' learning outcomes. And this is even more relevant within the field of science education.

As a consequence, this study aimed to evaluate the real impact of using ChatGPT as a virtual mentor on K-12 students learning science, within the frame of a blended-learning educational strategy, complementing the constructivist/connectivist presential learning with student-centered self-regulated location and time-independent cybergogy. More specifically, the AI was meant to assist teachers when they were unavailable, by virtually addressing students' doubts and homework correction in real-time, within a safe environment, through a personalized, meaningful and flexible learning experience that is independent of time and location, providing students with the support, advice and surveillance they might require.

First, the real competence of ChatGPT within K-12 chemistry and physics was systematically verified through a test designed for human students, paying special attention to encoding and the use of open-ended questions. ChatGPT provided answers with remarkable accuracy and insight to theoretical questions, and a clear, detailed and well-organized procedure to find the solution to problem-solving questions. The consistent and noticeable results obtained through this performance test (9.3/10 in 2023 -GPT-3.5 model-, and 9.7/10 in 2024 -GPT-4 model-) demonstrated ChatGPT provided trustworthy answers (97%) in the field of K-12 science education, despite some minor limitations that were duly discussed. These results suggested ChatGPT might be a reliable tool to help K-12 students learning science to reinforce theoretical knowledge and enhance their problem-solving skills with no time and location restrictions, when teachers are unavailable.

Furthermore, several aspects of the use of ChatGPT within the proposed pedagogical approach were discussed to be promoting meaningful interactions with students, from a theoretical perspective, considering current media landscape.

Then, the real impact of using ChatGPT as virtual mentor on K-12 students learning science, when teachers were unavailable, was assessed through a quasi-experimental analysis. The learning outcomes being monitored before and after the intervention were students' proficiency and students' perception of the AI as a useful educational tool.

On one hand, the grades of students belonging to the experimental group increased 30% after the intervention, three-times higher than that of the control group, which was manifested by 90% of

 doi:10.20944/preprints202409.1323.v1

students in the experimental group, independently of their level of proficiency. The class grades' average improved from low-medium to medium-high level of proficiency, with the only additional help of ChatGPT as an educational tool assisting students when teachers were unavailable, verifying the functional support the AI might offer to students.

On the other hand, students' perception on the AI as a useful educational tool was measured through a Likert scale, reaching an average of 4.05 (agree), with most students agreeing or strongly agreeing (70%). Besides, the study also revealed that students counting on no parent/tutor able to help them with their particular doubts when teachers were unavailable, felt reassurance counting on ChatGPT from now on, verifying the AI provided not only functional but also social/emotional support.

After a profound discussion, the study concluded that ChatGPT might be a useful educational tool able to furnish K-12 students learning science with the functional and social/emotional support they might require, democratizing a higher level of knowledge acquisition without parents/tutors help, also promoting students' autonomy, security and self-efficacy. These results probe ChatGPT's outstanding capacity (and vast potential) to assist teachers in their mentoring tasks, when they are unavailable, laying the foundations of virtual mentoring and paving the way for more discussion and future empirical studies allowing to get a more realistic perception of the AI impact on education.

Supplementary Materials: The following supporting information can be downloaded at: preprints.org. The document contains the study to assess ChatGPT's performance in the field of chemistry and physics for K-12 students, and the material used to evaluate ChatGPT's impact on real K-12 (15-16-years-old) students learning chemistry.

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

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the 1964 Declaration of Helsinki and its later amendments, the 1978 Belmont report, the EU Charter of Fundamental Rights (26/10/2012), the national and European ethical standards (European Network of Research Ethics Committees), and the EU General Data Protection Regulation (2016/679).

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

Data Availability Statement: The data presented in this study are unavailable because of privacy preservation reasons. Requests to access the data will require a certificate of approval by the I.E.S. Benaguasil, and the participants' consent.

Acknowledgments: The authors acknowledge students from I.E.S. Benaguasil participating within the study, and the I.E.S. for supporting the research. A.B. gratefully acknowledges financial support from Spanish national project No. PID2022-137857NA-I00. A.B. thanks MICINN for the Ramon y Cajal Fellowship (grant No. RYC2021-030880-I). F.J.D.-F. acknowledges the Next Generation EU program, Spanish National Research Council (Ayuda Margarita Salas), and Universitat Politècnica de València (PAID-06-23). E.P.-C acknowledges funding from Generalitat Valenciana (Grant No. SEJIGENT/2021/039) and AGENCIA ESTATAL DE INVESTIGACIÓN of Ministerio de Ciencia e Innovacion (PID2021-128442NA-I00).

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

References

1. Schwab, K. The Fourth Industrial Revolution. *Foreign Affairs*, 12 December **2015**.
2. Wang, Y., Ma, HS., Yang, JH., Wang, KS. Industry 4.0: a way from mass customization to mass personalization production. *Adv. Manuf.* **2017**, *5*, 311–320.
3. Schwab, K. The Fourth Industrial Revolution: what it means, how to respond. *World Economic Forum* **2016**. Available online: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/ (accessed on 19 February 2023).

4. Hilbert, M. and López, P. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science* **2011**, *332(6025)*, 60-65.
5. Esposito, M. World Economic Forum White Paper: Driving the Sustainability of Production Systems with Fourth Industrial Revolution Innovation. *World Economic Forum*, **2018**. Available online: https://www.researchgate.net/publication/322071988_World_Economic_Forum_White_Paper_Driving_the_Sustainability_of_Production_Systems_with_Fourth_Industrial_Revolution_Innovation (accessed on 20 February 2023).
6. Bondyopadhyay, P.K. In the beginning [junction transistor]. *Proceedings of the IEEE* **1998**, *86*, 63-77.
7. What are Industry 4.0, the Fourth Industrial Revolution, and 4IR? McKinsey, 17 August **2022**. Available online: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-are-industry-4-0-the-fourth-industrial-revolution-and-4ir (accessed on 20 February 2023).
8. Bai, C., Dallasega, P., Orzes, G., Sarkis, J. Industry 4.0 technologies assessment: A sustainability perspective. *International Journal of Production Economics* **2020**, 229, 107776.
9. Marr, B. Why Everyone Must Get Ready For The 4th Industrial Revolution. *Forbes* **2016**. Available online: https://www.forbes.com/sites/bernardmarr/2016/04/05/why-everyone-must-get-ready-for-4th-industrial-revolution/?sh=366e89503f90 (accessed on 22 February 2023).
10. Mudzar, N.M.B.M., Chew, K.W. Change in Labour Force Skillset for the Fourth Industrial Revolution: A Literature Review. *International Journal of Technology* **2022**, *13(5)*, 969-978.
11. Goldin, T.; Rauch, E.; Pacher, C.; Woschank, M. Reference Architecture for an Integrated and Synergetic Use of Digital Tools in Education 4.0. *Procedia Computer Science* **2022**, 200, 407-417.
12. Cónego, L., Pinto, R., Gonçalves, G. Education 4.0 and the Smart Manufacturing Paradigm: A Conceptual Gateway for Learning Factories. In *Smart and Sustainable Collaborative Networks 4.0*, Camarinha-Matos, L.M.; Boucher, X.; Afsarmanesh, H., Eds. PRO-VE 2021. IFIP Advances in Information and Communication Technology, vol 629. Springer, Cham.
13. Costan, E.; Gonzales, G.; Gonzales, R.; Enriquez, L.; Costan, F.; Suladay, D.; Atibing, N.M.; Aro, J.L.; Evangelista, S.S.; Maturan, F.; Selerio, E., Jr.; Ocampo, L. Education 4.0 in Developing Economies: A Systematic Literature Review of Implementation Barriers and Future Research Agenda. *Sustainability* **2021**, *13*, 12763.
14. González-Pérez, L.I.; Ramírez-Montoya, M.S. Components of Education 4.0 in 21st Century Skills Frameworks: Systematic Review. *Sustainability* **2022**, *14*, 1493.
15. Bonfield, C.A.; Salter, M.; Longmuir, A.; Benson, M.; Adachi, C. Transformation or evolution?: Education 4.0, teaching and learning in the digital age. *Higher Education Pedagogies for the 4th Industrial Revolution* **2020**, *5:1*, 223-246.
16. Miranda, J.; Navarrete, C.; Noguez, J.; Molina-Espinosa, J.M.; Ramírez-Montoya, M.S.; Navarro-Tuch, S.A.; Bustamante-Bello, M.R.; Rosas-Fernández, J.B.; Molina,A. The core components of education 4.0 in higher education: Three case studies in engineering education. *Computers & Electrical Engineering* **2021**, *93*, 107278.
17. Chiu,W.-K. Pedagogy of Emerging Technologies in Chemical Education during the Era of Digitalization and Artificial Intelligence: A Systematic Review. *Educ. Sci.* **2021**, *11*, 709.
18. Mhlanga, D., and Moloi, T. COVID-19 and the digital transformation of education: what are we learning on 4IR in South Africa? *Educ. Sci.* **2020**, *10*, 1–11.
19. Peterson, L.; Scharber, C.; Thuesen, A.; Baskin, K. A rapid response to COVID-19: one district's pivot from technology integration to distance learning. *Information and learning science* **2020**, *121(5-6)*, 461-469.
20. Guo, YJ; Chen, L; Guo, Yujuan; Chen, Li. Ninth International Conference of Educational Innovation through Technology (EITT) **2020**, 10-18.
21. Mogos, R., Bodea, C.N., Dascalu, M., & Lazarou, E., Trifan, L., Safonkina, O., Nemoianu, I. Technology enhanced learning for industry 4.0 engineering education. *Revue Roumaine des Sciences Techniques - Serie Électrotechnique et Énergétique* **2018**, *63*, 429-435.
22. Moraes, E.B., Kipper, L.M., Hackenhaar Kellermann, A.C., Austria, L., Leivas, P., Moraes, J.A.R. and Witczak, M. Integration of Industry 4.0 technologies with Education 4.0: advantages for improvements in learning. *Interactive Technology and Smart Education* **2022**, Vol. ahead-of-print, No. ahead-of-print.
23. Ciolacu, M.I., Tehrani, A.F., Binder, L., & Svasta, P. Education 4.0 - Artificial Intelligence Assisted Higher Education: Early recognition System with Machine Learning to support Students' Success. *IEEE 24th International Symposium for Design and Technology in Electronic Packaging (SIITME)* **2018**, 23-30.
24. Chen, Z., Zhang, J., Jiang, X., Hu, Z., Han, X., Xu, M., Savitha, & Vivekananda, G.N. Education 4.0 using artificial intelligence for students performance analysis. *Inteligencia Artificial* **2020**, *23 (66)*, 124-137
25. Tahiru, F. AI in Education: A Systematic Literature Review. *Journal of Cases on Information Technology* **2021**, *23(1)*, 1-20.
26. Miao, Fengchun & Holmes, Wayne & Huang, Ronghuai & Zhang, Hui. AI and education Guidance for policy-makers. UNESCO Publishing: Paris, France, 2021. https://unesdoc.unesco.org/ark:/48223/pf0000376709 (accessed on 8 Mars 2023).

27. Carbonell, J.R. AI in CAI: An artificial-intelligence approach to computer-assisted instruction. *IEEE transactions on man-machine systems* **1970**, *11(4)*, 190-202.
28. Psotka, J., Massey, L. D., & Mutter, S. A. (Eds.). (1988). Intelligent tutoring systems: Lessons learned. Lawrence Erlbaum Associates, Inc., New Jersey, United States.
29. S. Piramuthu. Knowledge-based web-enabled agents and intelligent tutoring systems. *IEEE Transactions on Education* **2005**, *48*, no. 4, 750-756.
30. Mousavinasab, E.; Zarifsanaiey, N.; Kalhori, S.R.N.; Rakhshan, M.; Keikha, L.; Saeedi, M.G. Intelligent tutoring systems: a systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments* **2021**, *29:1*, 142-163.
31. Alrakhawi, H.; Jamiat, N.; Abu-Naser, S. Intelligent tutoring systems in education: a systematic review of usage, tools, effects and evaluation. *Journal of Theoretical and Applied Information Technology* **2023**, *101*, 1205-1226.
32. Song, D.; Oh, E.Y.; Rice, M. Interacting with a conversational agent system for educational purposes in online courses. *10th International Conference on Human System Interactions (HSI)* **2017**, 78-82, Ulsan, Korea (South).
33. Shute, V. J., & Psotka, J. (1994). Intelligent Tutoring Systems: Past, Present, and Future. Human resources directorate manpower and personnel research division, 2-52.
34. VanLehn, K. "The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems". *Educational Psychologist* **2011**, *46 (4)*, 197–221.
35. Fernoaga, P.V.; Sandu, F.; Stelea, G.A.; Gavrila, C. Intelligent Education Assistant Powered by Chatbots. Conference proceedings of The 14th International Scientific Conference of eLearning and Software for Education (eLSE) **2018**, 376-383.
36. Hamam, D. (2021). The New Teacher Assistant: A Review of Chatbots' Use in Higher Education. In: Stephanidis, C., Antona, M., Ntoa, S. (eds) *HCI International 2021 - Posters. HCII 2021. Communications in Computer and Information Science* **2021**, *vol 1421*. Springer, Cham.
37. Satu, M.S.; Parvez, M.H. ; Shamim-Al-Mamun. Review of integrated applications with AIML based chatbot. *International Conference on Computer and Information Engineering (ICCIE)* 2015, Rajshahi, Bangladesh, 87-90.
38. Schwab, K. The Fourth Industrial Revolution: what it means, how to respond. World Economic Forum 2016. Available online: https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/ (accessed on 19 February 2023).
39. "The state of AI in 2023: Generative AI's breakout year" *McKinsey AI global survey* **2023**. https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-AIs-breakout-year#/ (accessed on 18 April 2024).
40. N. Maslej, L. Fattorini, R. Perrault, V. Parli, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J.C. Niebles, Y Shoham, R Wald, and J Clark. "The AI Index 2024 Annual Report," AI Index Steering Committee, Institute for Human-Centered AI, Stanford, **2024**. University, Stanford, CA, April 2024.
41. R. Lam et al. "Learning skillful medium-range global weather forecasting". *Science* **2023**, *382*, 1416-1421.
42. Merchant, A., Batzner, S., Schoenholz, S.S. et al. "Scaling deep learning for materials discovery". *Nature* **2023**, *624*, 80–85.
43. Boiko, D.A., MacKnight, R., Kline, B. et al. "Autonomous chemical research with large language models". *Nature* **2023**, *624*, 570–578.
44. Rudolph, J.; Tan, S.; Tan, S. ChatGPT: Bullshit spewer or the end of traditional assessments in higher education? *Journal of Applied Learning & Teaching* **2023**, *6*, 1.
45. Castelvecchi, D. Are ChatGPT and AlphaCode going to replace programmers? *Nature* **2022**. Published online December 8, 2022. Available from: https://doi.org/10.1038/d41586-022-04383-z (accessed on 13 Mars 2023).
46. Tung, L. ChatGPT can write code. Now researchers say it's good at fixing bugs, too. *ZDNET* **2023**. Archived from the original on February 3, 2023. Available from: https://www.zdnet.com/article/chatgpt-can-write-code-now-researchers-say-its-good-at-fixing-bugs-too/ (accessed on 13 Mars 2023).
47. Stokel-Walker, C. AI bot ChatGPT writes smart essays — should professors worry? *Nature* **2022**. Published online December 9, 2022. Available from: https://www.nature.com/articles/d41586-022-04397-7 (accessed on 13 Mars 2023).
48. Kung, TH.; Cheatham, M.; Medenilla, A.; Sillos, C.; De Leon, L.; Elepaño, C.; Madriaga, M.; Aggabao, R.; Diaz-Candido, G., Maningo, J.; Tseng, V. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health* **2023**, 2(2): e0000198.
49. Koe, C. ChatGPT shows us how to make music with ChatGPT. Published online January 27, 2023. Available from: https://musictech.com/news/gear/ways-to-use-chatgpt-for-music-making/ (accessed on 13 Mars 2023).

50. Zheng, Z.; Zhang, O.; Borgs, C.; Chayes, J. T.; Yaghi, O. M. "ChatGPT Chemistry Assistant for Text Mining and the Prediction of MOF Synthesis". *J. Am. Chem. Soc.* **2023**, *145 (32)*, 18048– 18062.
51. Pradhan, T. et al. "The Future of ChatGPT in Medicinal Chemistry: Harnessing AI for Accelerated Drug Discovery." *Chemistry Select* **2024**, *9*, 13, e202304359.
52. Zhang W, Wang Q, Kong X, Xiong J, Ni S, Cao D, et al. "Fine-tuning Large Language Models for Chemical Text Mining". *ChemRxiv*. **2024.**
53. OpenAI (2023). https://arxiv.org/pdf/2303.08774.pdf (accessed on 21 Mars 2023).
54. Roose, K. The Brilliance and Weirdness of ChatGPT. *The New York Times*. Published online December 5, **2022**. Available from: https://www.nytimes.com/2022/12/05/technology/chatgpt-ai-twitter.html (accessed on 14 Mars 2023).
55. Sanders, N.E.; Schneier, B. Opinion | How ChatGPT Hijacks Democracy. *The New York Times.* Published online on January 15, **2023**. Available from: https://archive.is/Cyaac (accessed on 14 Mars 2023).
56. García-Peñalvo, F. J. La percepción de la Inteligencia Artificial en contextos educativos tras el lanzamiento de ChatGPT: disrupción o pánico. *Education in the Knowledge Society (EKS)* **2023**, 24, e31279.
57. Chomsky, N.; Roberts, I.; Watumull, J. Opinion | Noam Chomsky: The False Promise of ChatGPT. *The New York Times*. Published online on March 12, **2023**. Available from: https://archive.is/SM77M (accessed on 14 Mars 2023).
58. Lo, C.K. What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Educ. Sci.* **2023**, *13*, 410.
59. Terwiesch, C. Would ChatGPT get a Wharton MBA? A prediction based on its performance in the operations management course. Mack Institute for Innovation Management at the Wharton School, University of Pennsylvania; 2023. Available from: https://mackinstitute.wharton.upenn.edu/wp-content/uploads/2023/01/Christian-Terwiesch-Chat-GTP.pdf (accessed on 13 Mars 2023).
60. Mogali, S.R. Initial impressions of ChatGPT for anatomy education. *Anat Sci Educ.* **2023** (07 February).
61. Wu, R., Yu, Z. Do AI chatbots improve students learning outcomes? Evidence from a meta-analysis. *Br J Educ Technol.* **2023**, *00*, 1–24.
62. Rospigliosi, P. Artificial intelligence in teaching and learning: what questions should we ask of ChatGPT? *Interactive Learning Environments* **2023**, *31:1*, 1-3.
63. Pavlik, J. V. Collaborating With ChatGPT: Considering the Implications of Generative Artificial Intelligence for Journalism and Media Education. *Journalism & Mass Communication Educator* **2023**, *78(1)*, 84–93.
64. Jeon, J.; Lee, S. Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Educ Inf Technol* **2023**.
65. Luan, L.; Lin, X.; Li, W. Exploring the Cognitive Dynamics of Artificial Intelligence in the Post-COVID-19 and Learning 3.0 Era: A Case Study of ChatGPT. arXiv:2302.04818 **2023**.
66. Rahman, M.M.; Watanobe, Y. ChatGPT for Education and Research: Opportunities, Threats, and Strategies. *Appl. Sci.* **2023**, 13, 5783.
67. Kamil Malinka, Martin Peresíni, Anton Firc, Ondrej Hujnák, Filip Janus. On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree? *ITiCSE 2023: Proceedings of the 2023 Conference on Innovation and Technology in Computer Science Education V. 1,* **2023**, 47–53.
68. Zhai, X. ChatGPT for Next Generation Science Learning (January 20, **2023**). Available at *SSRN*: https://ssrn.com/abstract=4331313 or http://dx.doi.org/10.2139/ssrn.4331313
69. Wollny Sebastian, Schneider Jan, Di Mitri Daniele, Weidlich Joshua, Rittberger Marc, Drachsler Hendrik. Are We There Yet? - A Systematic Literature Review on Chatbots in Education. *Frontiers in Artificial Intelligence* **2021**, *4*.
70. Cooper, G. Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence. *J Sci Educ Technol* **2023**, *32*, 444–452.
71. Dos Santos, R.P. Enhancing Chemistry Learning with ChatGPT, Bing Chat, Bard, and Claude as Agents-to-Think-With: A Comparative Case Study. *arXiv:2311.00709* **2023**.
72. Schulze Balhorn, L., Weber, J.M., Buijsman, S. et al. Empirical assessment of ChatGPT's answering capabilities in natural science and engineering. *Sci Rep* **2024**, *14*, 4998.
73. Su (苏嘉红), J.; Yang (杨伟鹏), W. Unlocking the Power of ChatGPT: A Framework for Applying Generative AI in Education. ECNU Review of Education 2023, 0(0).
74. L. Mercadé, F.J. Díaz-Fernández, M. Sinusia Lozano, J. Navarro-Arenas, M. Gómez, E. Pinilla-Cienfuegos, D. Ortiz de Zárate, V.J. Gómez Hernández, A. Díaz-Rubio. Research mapping in the teaching environment: tools based on network visualizations for a dynamic literature review. *INTED2023 Proceedings* **2023**, 3916-3922.
75. L. Mercadé, D. Ortiz de Zárate, A. Barreda, E. Pinilla-Cienfuegos. *INTED2023 Proceedings* **2023**, 6175-6179.
76. A. Barreda, B. García-Cámara, D. Ortiz de Zárate Díaz, E. Pinilla-Cienfuegos, L. Mercadé. *INTED2023 Proceedings* **2023**, 2547-2554.
77. Bizami, N.A.; Tasir, Z.; Kew, S.N. Innovative pedagogical principles and technological tools capabilities for immersive blended learning: a systematic literature review. *Educ. Inf. Technol.* **2023**, *28*, 1373–1425.

78. Chen, C.K.; Huang, N.T.N.; Hwang, G.J. Findings and implications of flipped science learning research: A review of journal publications. *Interactive Learning Environments* **2022**, *30:5*, 949-966.
79. Stahl, B.C.; Eke, D. The ethics of ChatGPT – Exploring the ethical issues of an emerging technology. *International Journal of Information Management* **2024**, *74*, 102700.
80. Wu, X.; Duan, R.; Ni, J. Unveiling security, privacy, and ethical concerns of ChatGPT. *Journal of Information and Intelligence* **2024,** *2, 2*, 102-115.
81. Peng, L., & Zhao, B. Navigating the ethical landscape behind ChatGPT. *Big Data & Society* **2024**, 11(1).
82. Zhou, J.; Müller, H.; Holzinger, A.; Chen, F. Ethical ChatGPT: Concerns, Challenges, and Commandments. *arXiv* **2023**. https://arxiv.org/pdf/2305.10646
83. 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.
84. Ferrara, E. Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models. *arXiv* **2023**. https://arxiv.org/pdf/2304.03738.pdf
85. Jenkinson, J. Measuring the Effectiveness of Educational Technology: What are we Attempting to Measure? *Electronic Journal of e-Learning* **2009**, 7, 3, 273 – 280.
86. Gilson A, Safranek CW, Huang T, Socrates V, Chi L, Taylor RA, 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**, Feb 8, 9, e45312.
87. Turing, Alan. Computing Machinery and Intelligence. *Mind, LIX* **1950**, *(236)*, 433–460.
88. Boone, H. N., & Boone, D. A. Analysing Likert data. *Journal of extension* **2012**, *50(2)*, 1–5.
89. Wu, S. and Wang, F. Artificial intelligence-based simulation research on the flipped classroom mode of listening and speaking teaching for English majors. *Mobile Information Systems* **2021**, Article ID 4344244.
90. Klos MC, Escoredo M, Joerin A, Lemos VN, Rauws M, Bunge EL. Artificial Intelligence-Based Chatbot for Anxiety and Depression in University Students: Pilot Randomized Controlled Trial. *JMIR Form Res.* **2021**, 5(8), e20678.
91. Mishra P, Singh U, Pandey CM, Mishra P, Pandey G. Application of student's t-test, analysis of variance, and covariance. *Ann Card Anaesth.* **2019**, 22(4), 407-411.
92. Hsu, M. H., Chen, P. S., & Yu, C. S. Proposing a task-oriented chatbot system for EFL learners speaking practice. *Interactive Learning Environments* **2021**, 1–12.
93. Ilgaz H B, Çelik Z. The Significance of Artificial Intelligence Platforms in Anatomy Education: An Experience With ChatGPT and Google Bard. *Cureus* **2023**, *15(9)*: e45301.
94. Litt, E., Zhao, S., Kraut, R., & Burke, M. What Are Meaningful Social Interactions in Today's Media Landscape? A Cross-Cultural Survey. *Social Media + Society* **2020**, 6(3).
95. Cooper, H.; Okamura, L.; Gurka, V. Social activity and subjective well-being. *Personality and Individual Differences* **1992**, *13, 5*, 573-583.
96. Hilvert-Bruce Z., Neill J. T., Sjöblom M., Hamari J. Social motivations of live-streaming viewer engagement on Twitch. *Computers in Human Behavior* **2018**, *84*, 58–67.
97. Offer S. Family time activities and adolescents' emotional well-being. *Journal of Marriage and Family* **2013**, *75(1),* 26–41.
98. Gonzales A. L. Text-based communication influences self-esteem more than face-to-face or cellphone communication. *Computers in Human Behavior* **2014**, *39*, 197–203.
99. Brennan S. E. The grounding problem in conversations with and through computers. In Fussell S. R., Kreuz R. J. (Eds.), *Social and cognitive approaches to interpersonal communication* **1998** (pp. 201–225). Lawrence Erlbaum.
100. Boothby E. J., Clark M. S., Bargh J. A. Shared experiences are amplified. *Psychological Science* **2014**, *25(12),* 2209–2216.
101. Maslow, A. H. Preface to motivation theory. *Psychosomatic medicine* **1943**, *5(1),* 85-92.
102. Deci, E. L.; Ryan, R. M. Autonomy and need satisfaction in close relationships: Relationships motivation theory. *Human motivation and interpersonal relationships: Theory, research, and applications* **2014**, 53-73.
103. Burleson, B. R. The experience and effects of emotional support: What the study of cultural and gender differences can tell us about close relationships, emotion and interpersonal communication. *Personal Relationships*, **2003**, *10(1),* 1–23.
104. Cook, T. D., & Campbell, D. T. **1979**. Quasi-experimentation: Design & analysis issues for field settings (1st ed.). Chicago: Rand McNally.
105. Caspersen, J.; Smeby, J.C.; Aamodt, P.O.. Measuring learning outcomes. *Eur J Educ.* **2017**, 52, 20–30.
106. The importance of grades. **2017**. Urban Education Institute. University of Chicago. https://uei.uchicago.edu/sites/default/files/documents/UEI%202017%20New%20Knowledge%20-%20The%20Importance%20of%20Grades.pdf (accessed on 13 July 2024).
107. J. Moon, R. Yang, S. Cha and S. B. Kim. ChatGPT vs Mentor : Programming Language Learning Assistance System for Beginners. *2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS)*, Penang, Malaysia, **2023**, pp. 106-110.

108. Kim, N. Y. A study on the use of artificial intelligence chatbots for improving English grammar skills. *Journal of Digital Convergence* **2019**, *17(8)*, 37–46.
109. Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. Educational AI chatbots for content and language integrated learning. *Applied Sciences* **2022**, *12(7)*, Article 7.
110. Hwang, W. Y., Guo, B. C., Hoang, A., Chang, C. C., & Wu, N. T. Facilitating authentic contextual EFL speaking and conversation with smart mechanisms and investigating its influence on learning achievements. *Computer Assisted Language Learning* **2022**, 1–27.
111. Garzón, J., & Acevedo, J. Meta-analysis of the impact of augmented reality on students' learning gains. *Educational Research Review* **2019**, *27*, 244–260.
112. Jeon, J. Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning* **2022**, 1–26.
113. Watson, S., & Romic, J. ChatGPT and the entangled evolution of society, education, and technology: A systems theory perspective. *European Educational Research Journal* **2024**, *0(0)*.
114. Anderson, P.W. More is different. *Science* **1972**, *177*, 393-396.