Analysing the Impact of Artificial Intelligence and Computational Sciences on Student Performance: Systematic Review and Meta-analysis

Inmaculada García-Martínez¹, José María Fernández-Batanero², José Fernández-Cerero² and Samuel P. León³

¹Department of Didactics and School Organization, University of Granada, Spain
²Department of Didactics and School Organization, University of Seville, Spain
³Department of Pedagogy, University of Jaén, Spain

ABSTRACT

Artificial intelligence (AI) and computational sciences have aroused a growing interest in education. Despite its relatively recent history, AI is increasingly being introduced into the classroom through different modalities, with the aim of improving student achievement. Thus, the purpose of the research is to analyse, quantitatively and qualitatively, the impact of AI components and computational sciences on student performance. For this purpose, a systematic review and meta-analysis have been carried out in WOS and Scopus databases. After applying the inclusion and exclusion criteria, the sample was set at 25 articles. The results support the positive impact that AI and computational sciences have on student performance, finding a rise in their attitude towards learning and their motivation, especially in the STEM (Science, Technology, Engineering, and Mathematics) areas. Despite the multiple benefits provided, the implementation of these technologies in instructional processes involves a great educational and ethical challenge for teachers in relation to their design and implementation, which requires further analysis from the educational research. These findings are consistent at all educational stages.

Keywords ARTIFICIAL INTELLIGENCE, LEARNING ACHIEVEMENT, EDUCATION, TEACHING METHODS, EDUCATIONAL IMPROVEMENT

1 INTRODUCTION

Nowadays, society is becoming increasingly oriented towards a massive process of technologicalisation in all spheres (political, economic, educational, social, etc.). This trend of adapting to new technological interaction communities has created a variety of technologies that allow communication with the user, called “virtual assistants”, which use computer algorithms to emulate human intelligence so that users have the feeling that they are interacting with another person. This concept is known as “artificial intelligence” (AI) (Ocaña,
Valenzuela, & Garro, 2019; Yang, Zhuang, & Pan, 2021). In educational environments, AI has taken special interest, given the high possibilities of communication that are established between teachers and students when using virtual information assistants, since from its execution there is a simulation of responses that approach a human conversation and, as the tool is used, the interaction with the user is learned and recognised intuitively. Nevertheless, in the current global context of the technological revolution, there are human qualities that cannot yet be reproduced by AI, such as creativity or the ability to produce new ideas or to improvise and constantly evolve. In this regard, current trends in AI and computational science are moving towards human-centred AI, that considers people’s characteristics and contexts to reduce the biases that may be associated with the management and processing of the algorithms that support it (Yang et al., 2021). According to UNESCO (2019), the link between AI and education consists of three areas: learning with AI (using AI tools in the classroom), learning about AI (its technologies and techniques) and preparing for AI (enabling all citizens to understand the potential impact of AI on human life). It also believes that AI has the potential to address some of the biggest challenges facing education today, namely, to develop innovative teaching and learning practices guided by the fundamental principles of inclusion and equity while helping to accelerate progress towards SDG 4.

Thus, AI has a strong potential to accelerate the process of realisation and development of the global goals around education by reducing barriers to access to learning, by automating management processes, and by optimising methods to improve student performance and, as a result, learning outcomes (Moreno, 2019).

In short, AI as a technological tool applied to education may contribute innovative methods and ways with the use of ICT that improve the teaching and learning process from the perspective of the student and the teacher.

1.1 Artificial Intelligence in the Educational Context

The educational field is constantly changing and adapting to new generations and their educational needs (Halili, 2019). If we consider that all these developments go in parallel with technological advances, we can say that the speed at which education is being updated may be the fastest ever (Harrison, 1986). Although these advances are studied from their many forms, they all have a common factor: AI.

To address the concept of AI, as discussed above, it must be understood to include any resource or machine that carries out human work. Humans create these machines to mechanise the tasks they perform every day, with the purpose of accomplishing more in less time. Popenici and Kerr (2017) define it as computing systems that can engage in similar processes to the human ones, such as learning, adapting, synthesising, self-correcting and using data for complex processing tasks. More specifically in education, educational artificial intelligence (EAI) refers to the use of AI to support personalised and automated feedback and guidance in the educational field Song and Wang (2020).

The growing demand in education in recent years has given rise to a thriving new field of research which integrates AI and education, which has resulted in an expansion
of the existing literature on EAI. Furthermore, and updating this terminology, EAI is related to different fields, such as robotics (Jawaid et al., 2020), applications for smart devices (Petko, Schmid, Müller, & Hielscher, 2019), electronic devices (Pyörälä et al., 2019), e-learning (Reister & Blanchard, 2020; Singer-Brodowski, Brock, Etzkorn, & Otte, 2019), or virtual (VR) and augmented reality (AR) (Bower, Dewitt, & Lai, 2020; Kavanagh, Luxton-Reilly, Wuensche, & Plimmer, 2017), intelligent conversational software agents (chatbot) (Schachner, Keller, & Wangenheim, 2020), virtual assistants (Jee, 2019), and online platforms for self-learning (Moreno, 2019). All these areas have a common goal: to learn, to teach and to solve problems (Baker, Smith, & Anissa, 2019).

Several studies have highlighted the different current trends in EAI (Roll & Wylie, 2016); thus, we can highlight the role of AI in special education (Guilherme, 2017), such as the transformative collaboration between teachers and students (Drigas & Ioannidou, 2013), the global advance and trends of various intelligent tutoring systems (Han, Zhao, Jiang, Oubibi, & Hu, 2019), the potential of EAI in higher education (Crompton, Bernacki, & Greene, 2020), etc. Hence, the importance to contribute to the improvement of academic performance by optimising instructional processes and reducing students’ learning difficulties.

In our study, we understand academic performance as a measure of the student’s abilities, which expresses what the student has learned throughout the learning process. It also implies the student’s ability to respond to educational stimuli (Martínez, Karanik, Giovannini, & Pinto, 2015). Academic performance is affected by a multiplicity of heterogeneous factors (internal and external) that condition student performance, including aptitude and motivation (Castrillón, Sarache, & Herrera, 2020).

Research in the field of EAI and academic performance is a key area in education, since through its study the effectiveness of all these intelligent resources in the teaching and learning processes can be evaluated.

Considering all the above, this study aims to conduct an empirical analysis of the evidence found within the EAI literature. Despite previous systematic reviews on the inclusion of AI in education (Hooshyar, Yousefi, & Lim, 2019; Roll & Wylie, 2016; Song & Wang, 2020; Zawacki-Richter, Marín, Bond, & Gouverneur, 2019), the effect size on performance has not yet been calculated, quantified or meta-analysed. Thus, the main objective of the research is to analyse the impact of AI components and computational sciences on student performance. Focusing on EAI, this study attempts to address the following research questions:

- Does EAI improve student performance?
- What effects does EAI have on students?
- What type of AI and computational science is the most common in the educational field?
- Is EAI effective at all educational stages?
2 METHODS

2.1 Search Procedures

This study follows the guidelines of the PRISMA Statement (Page et al., 2021). The search was conducted on Web of Science (WOS) and Scopus databases, due to their prestige and scope in education research. For this purpose, the following search equation was introduced: SUBJECT: (AI-based education) OR SUBJECT: (artificial intelligence in education) AND SUBJECT: (e-learning) OR SUBJECT: (automated tutor) OR SUBJECT: (intelligence agent) OR SUBJECT: (simulation) OR SUBJECT: (artificial agent) AND SUBJECT: (student) AND SUBJECT: (experimental design) OR SUBJECT: (quasi-experimental design).

The following inclusion criteria were adopted: (a) only articles, excluding the so-called grey literature; (b) whose language was English, as it is the internationally recognised language in the scientific field; (c) published between 2010 and 2020. In the search process, the volatility of educational technology was assumed, so it was decided to focus on empirical research published in the last 10 years; (d) belonged to the area of “Education Educational research or Education Scientific disciplines” in the case of WOS, and the research areas “social sciences, art and humanities, psychology and neuroscience” in Scopus. From the application of the initial inclusion criteria, the search was set at 480 manuscripts: 455 WOS and 25 in Scopus. There were 87 duplicates, so the initial sample was set at 393.

A first screening of the papers, resulting from the initial search, was conducted by examining the titles and abstracts. This phase was carried out by two of the authors. Unpublished dissertations, reviews and meta-analyses were excluded at this stage. This first screening resulted in a total of 41 articles that were fully read independently by the authors to verify whether the inclusion criteria were met. Of the 41 articles, only 18 fully achieved the inclusion criteria. A top-down search of articles citing or cited by those two articles was then performed to identify possible additional studies. This allowed the inclusion of five new articles. Finally, we looked for previous systematic reviews on AI, or some of its components, and analysed all the papers cited in those reviews for possible inclusion, considering two more studies in this last search. Thus, the final sample of articles reviewed for inclusion in this study was 25 articles. Figure 1 summarises the search process in a PRISMA flowchart.

2.2 Selection Criteria

The studies analysed must meet the following criteria to be included in this review: (a) the objective of the study was to measure quantitatively the impact of EAI, or one of its components, on the academic development of students; (b) the studies had to follow a pre-post design with control groups; (c) the dependent variable had to be related to academic performance; (d) the sample had to be made up of students.

2.3 Data Extraction and Coding

The review process of the manuscripts to reach the sample was carried out independently by two of the authors who signed the article. For this purpose, the data from the articles
were analysed in an Excel file, according to the parameters provided by the inclusion and exclusion criteria, where a cell was added in order to determine whether or not they were included in the review and the reason for them. After completing each phase of the review (reading the title and abstract and the full text), the results obtained were contrasted.

The methodological quality of the selected articles was also assessed using Joanna Briggs’ checklist (JBI), where it was examined by means of critical and independent review (eleven-point checklist) (Aromataris & Munn, 2020). For the internal assessment of the quality of the proposed study, the checklist was evaluated blindly by two researchers external to the research, with the aim of avoiding assessment bias by the authors themselves.

For the first complete reading made by the authors in the first search, the degree of agreement of inclusion of the papers was 96%. Disagreements were resolved by discussion and consensus between the two researchers until there was a 100% agreement. In the second search, the degree of agreement was 100%. The studies included were finally analysed in depth by the authors. All the relevant information of each article has been coded for the analysis and discussion procedure using a database, whose information has been interpolated into the figures and tables. Table 1 shows a summary of the detailed analysis of the 25 articles included in the systematic review.
<table>
<thead>
<tr>
<th>Autors and Year</th>
<th>Journal</th>
<th>Country</th>
<th>N</th>
<th>Stage</th>
<th>Type of intervenor</th>
<th>Intervention duration</th>
<th>DV</th>
<th>Tests used to measure DV</th>
<th>Results</th>
<th>Type of IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>J. L. Anderson and Barnett (2013)</td>
<td>Journal of Science Education and Technology</td>
<td>USA</td>
<td>91</td>
<td>C</td>
<td>CG = 32, EG = 59</td>
<td>Supercharged!</td>
<td>7 class periods</td>
<td>SA; CT</td>
<td>EG has better SA.</td>
<td>Ap</td>
</tr>
<tr>
<td>Barbalios, Ioannidou, Tzionas, and Paraskeuopoulos (2013)</td>
<td>Computer &amp; Education</td>
<td>Greece</td>
<td>24</td>
<td>C</td>
<td>3D virtual reality modelling language</td>
<td>-</td>
<td>SA; CT</td>
<td>EG has better SA.</td>
<td>VR</td>
<td></td>
</tr>
<tr>
<td>Ibáñez, Serio, Villarán, and Kloos (2014)</td>
<td>Computer &amp; Education</td>
<td>Spain</td>
<td>64</td>
<td>C</td>
<td>AR-based methodology vs. Web-based methodology</td>
<td>5 sessions</td>
<td>SA; P</td>
<td>The AR-based application was more effective than the web-based application.</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td>Borntnik, Stozhko, Pervukhina, Tchernysheva, and Belysheva (1968)</td>
<td>Research in Learning Technology</td>
<td>Russia</td>
<td>50</td>
<td>U</td>
<td>Virtual chemistry lab combined with classroom vs. Traditional teaching</td>
<td>-</td>
<td>SA</td>
<td>R of EG and CG; CT; SP</td>
<td>Sim</td>
<td></td>
</tr>
<tr>
<td>Chin et al. (2010)</td>
<td>Educational Technology Research and Development</td>
<td>USA</td>
<td>Include 2 studies: 1) EG = 28, CG = 30 2) n = 104</td>
<td>C</td>
<td>Teachable agents application</td>
<td>1) 3 weeks 2) 15 weeks</td>
<td>SA; CT</td>
<td>EG has better SA.</td>
<td>Ap</td>
<td></td>
</tr>
<tr>
<td>Givelek, Ucar, Ustunel, and Aydin (2014)</td>
<td>Eurasia Journal of Mathematics Science and Technology Education</td>
<td>Turkey</td>
<td>EG = 106, CG = 109</td>
<td>C</td>
<td>Immersive virtual reality environment (VRE)</td>
<td>2 weeks</td>
<td>SA; A</td>
<td>EG has better SA.</td>
<td>VR</td>
<td></td>
</tr>
<tr>
<td>Yelamarthi and Drake (2014)</td>
<td>IEEE Transactions on Education</td>
<td>USA</td>
<td>EG = 17, CG = 24</td>
<td>U</td>
<td>Flipped course</td>
<td>Twice a week for 75 min over the 16 weeks</td>
<td>SA; CT</td>
<td>Significant improvements in EG students' performance and their perceptions of their learning experience.</td>
<td>Sim</td>
<td></td>
</tr>
</tbody>
</table>

*Continued on next page*
### Table 1 continued

<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Journal</th>
<th>Country</th>
<th>N</th>
<th>Stage</th>
<th>Type of interventor</th>
<th>Intervention duration</th>
<th>DV</th>
<th>Tests used to measure DV</th>
<th>Results</th>
<th>Type of IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Olde, De Jong, and Gijlers (2013)</td>
<td>Educational Technology &amp; Society</td>
<td>Netherla</td>
<td>EG = 21</td>
<td>TVS</td>
<td>Look-experiment-design (LED) by computer simulation</td>
<td>3 two-hour sessions</td>
<td>SA</td>
<td>CT</td>
<td>EG has better SA.</td>
<td>Sim</td>
</tr>
<tr>
<td>Dickerson and Clark (2018)</td>
<td>Computer Applications in Engineering Education</td>
<td>USA</td>
<td>CG = 31</td>
<td>U</td>
<td>SPICE simulation</td>
<td>12 weeks</td>
<td>SA</td>
<td>I, CT</td>
<td>EG has better and deeper SA.</td>
<td>Sim</td>
</tr>
<tr>
<td>Fang and Guo (2016)</td>
<td>Journal of Computer Assisted Learning</td>
<td></td>
<td>U</td>
<td>EG = 77</td>
<td>Computer simulation and animation (CSA)</td>
<td>12 months</td>
<td>SA</td>
<td>CT</td>
<td>Better results with CSA, although they point out that the teacher’s role cannot be replaced.</td>
<td>Ap</td>
</tr>
<tr>
<td>Fidan and Tuncel (2019)</td>
<td>Computer &amp; Education</td>
<td>Turkey</td>
<td>EG1 = 30</td>
<td>AR based PBL with FenAR software</td>
<td>11 weeks</td>
<td>SA</td>
<td>CT; AS</td>
<td>The experimental results indicated that the integration of AR into the activities of PBL increased SA and AS.</td>
<td>AR</td>
<td></td>
</tr>
<tr>
<td>Jiménez, Bravo, and Bacca (2010)</td>
<td>Computer Applications in Engineering Education</td>
<td>Mexico</td>
<td>EG = 31</td>
<td>U</td>
<td>Web-based gamified software</td>
<td>20 days</td>
<td>SA</td>
<td>CT; MQ</td>
<td>EG scored better on the content test and displayed higher motivation than CG.</td>
<td>BL</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Journal</th>
<th>Country</th>
<th>N</th>
<th>Stage</th>
<th>Type of interventor</th>
<th>Intervention duration</th>
<th>DV</th>
<th>Tests used to measure DV</th>
<th>Results</th>
<th>Type of IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Madathil et al. (2017)</td>
<td>Computers in Education Journal</td>
<td>USA</td>
<td>165</td>
<td>U</td>
<td>Instructional model with VR integrated</td>
<td>1 hour</td>
<td>SA, Sa, U; Pe</td>
<td>CT</td>
<td>Although no significant learning differences were found between the conditions, students’ perceptions of their learning showed significant improvements in the VR and control groups over the photo-based group. VR improves the learning experience.</td>
<td>VR</td>
</tr>
<tr>
<td>Pellias and Vosinakis (2018)</td>
<td>Education and Information Technologies</td>
<td>Greece</td>
<td>50</td>
<td>Scratch and OpenSim with Scratch4SL</td>
<td>4-weeks period with 6 sessions</td>
<td>SA</td>
<td>CT</td>
<td>EG performed better on problem-solving measures and algorithmic thinking.</td>
<td>Sim</td>
<td></td>
</tr>
<tr>
<td>Stieff (2011)</td>
<td>Journal of Research in Science Teaching</td>
<td>USA</td>
<td>460</td>
<td></td>
<td>Connected Chemistry: Discovering Matter!</td>
<td>180 minutes with different distributions in each case: 4x45; 2x90...</td>
<td>RC; SA</td>
<td>CT</td>
<td>Connected Chemistry class yield only small to modest improvements in SA but higher RC.</td>
<td>Sim</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Journal</th>
<th>Country</th>
<th>N</th>
<th>Stage</th>
<th>Type of interventor</th>
<th>Intervention duration</th>
<th>DV</th>
<th>Tests used to measure DV</th>
<th>Results</th>
<th>Type of IA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zacharia and Olympiou (2011)</td>
<td>Learning and Instruction</td>
<td>Cyprus</td>
<td>234</td>
<td>U</td>
<td>Traditional vs. Physical manipulative experimentation vs. Virtual manipulative experimentation (VME)</td>
<td>2 semesters Duration: 30.5 hours</td>
<td>SA</td>
<td>CT</td>
<td>Better results in the experimental groups.</td>
<td>VR</td>
</tr>
<tr>
<td>J. Anderson and Barnett (2011)</td>
<td>Journal of Science Education and Technology</td>
<td>USA</td>
<td>CG = 65, EG = 71</td>
<td>U</td>
<td>Labs with and without Supercharged!</td>
<td>2 labs of 2 h each</td>
<td>SA</td>
<td>CT; LN</td>
<td>Video games can lead to positive learning outcomes and support student scientific understanding.</td>
<td>Ap</td>
</tr>
<tr>
<td>Bozkurt and Ilik (2010)</td>
<td>Procedia Social and Behavioral Sciences</td>
<td>Turkey</td>
<td>152</td>
<td>U</td>
<td>Computer simulations (E) vs. Traditional (C)</td>
<td>1 semester</td>
<td>SA; B</td>
<td>CT</td>
<td>Groups who study with computer simulations are more successful than those who study with traditional methods.</td>
<td>Sim</td>
</tr>
<tr>
<td>Pareto (2014)</td>
<td>International Artificial Intelligence in Education Society</td>
<td>Sweden</td>
<td>443</td>
<td>P</td>
<td>Teachable agents application</td>
<td>3 months</td>
<td>SA</td>
<td>CT</td>
<td>Teachable agents in educational games can help achieve deeper levels of learning and motivational power.</td>
<td>Ap</td>
</tr>
</tbody>
</table>

Continued on next page
<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Journal</th>
<th>Country</th>
<th>N</th>
<th>Stage</th>
<th>Type of interventor</th>
<th>Intervention duration</th>
<th>DV</th>
<th>Tests used to measure DV</th>
<th>Results</th>
<th>Type of IA</th>
</tr>
</thead>
</table>
| Riess and Mischo (2010) | International Journal of Science Education | Germany | CG = 84  
EG1 = 115  
EG2 = 112  
EG3 = 113 | S  | Four conditions: computer-simulated forest game only vs. Teaching unit on systems thinking vs. Combination of computer-simulated forest game and teaching unit on systems thinking vs. “Traditionally” lessons | 26 lessons | SA | CT | Significant increase in systems thinking can only be seen in the experimental condition “simulation and lessons”, whereas the other experimental conditions show either no increase or only a tendential one. | Sim |
| Shegog et al. (2012) | Research in Science Education | USA | 44 treatment = 23; comparison = 21 | S | Virtual lab | 1 day | SA; ATS, ATS | EG increased their procedural and declarative knowledge, became more positive toward using computers for learning but did not affect attitudes toward science. | Sim |
CG2 = 30  
EG = 30 | S  | Quasi-experimental method | 6 weeks | SA | CT; UO | Virtual laboratories are at least as effective as physical ones. | Ap |
CG= 42 | U | ePULab tool | 4 hours | SA | CT | Students using ePULab gave significantly better learning acquisition scores than those following traditional lecture-style classes. | AVT |

*Continued on next page*
<table>
<thead>
<tr>
<th>Authors and Year</th>
<th>Journal</th>
<th>Country</th>
<th>N</th>
<th>Stage</th>
<th>Type of interventor</th>
<th>Intervention duration</th>
<th>DV</th>
<th>Tests used to measure DV</th>
<th>Results</th>
<th>Type of IA</th>
</tr>
</thead>
</table>

Note. DV: Dependent Variable; CG: Control Group; EG: Experimental Group; P: Primary Education; S: Secondary Education; U: University; TVS: Technical Vocational School; SA: Student Achievement; ATS: Attitude to Science; B: Beliefs; RC: Representation Competence; ATC: Attitudes Toward Computers for Learning; SP: Student Portfolios R: Reports; PQ: Perception Questionnaire; CT: Content Tests; I: Interviews; MQ: Motivation questionnaire; Sa: Satisfaction; U: Usability; PE: Perceived Engagement; LN: Laboratory Notebooks; UO: Unstructured Observations; VR: Virtual Reality; AR: Augmented Reality; Ap: Application; Sim: Simulations; BL: B-learning; AVT: Artificial Vision Techniques.
2.4 Computation of Effect Sizes and Statistical Analyses

To analyse quantitatively the impact of EAI-based interventions on academic achievement, the effects size of the results of each of the included studies was estimated. One of the inclusion criteria was that the studies employed a control group design and pre-post measures, so the standardized mean change difference score recommended by Morris (2008) was used to estimate effect size. However, although all studies met the design criteria, 11 of them did not provide information on the pre-test measure. For these studies, the effect size was estimated by means of the standardized mean difference of post-test scores, and a second analysis was carried out with them.

Standardized mean change difference scores ($g_\Delta$) were calculated with equations 8-10 of Morris (2008); and the variance of the effects, with equation 25. For the analysis focused only on the post-test, the effect size ($g$) was computed using equations 4.18 and 4.19 of Borenstein, Hedges, Higgins, and Rothstein (2009), including the correction factor $J$, calculated with equations 4.22 and 4.23. As for the variance of $g_p$, equations 4.20 and 4.24 of Borenstein et al. (2009) were used. Since some of the included studies offered more than one dependent variable to analyze the statistical dependence of the effect sizes within the same study, the analysis with a multi-level random-effects model using the rma.mv function of the metafor package for R was adjusted (Viechtbauer, 2010).

3 RESULTS

3.1 Characterization of the Studies Examined

The studies in this systematic review share the objective related to assess the impact of AI-based models and computational sciences on student achievement. The wide range of research about AI and computational sciences has led to the inclusion of empirical research on different types developed at different educational stages around the world (see Table 1).

Among the studies conducted in primary education, the research of Chin et al. (2010), Barbalios et al. (2013) and Pareto (2014) should be mentioned. The first one addresses teachable agents with K-12 students, and its purpose was to facilitate student learning through concept mapping with an AI tool. Their findings reported that the use of this tool promotes the consolidation and acquisition of future learning in the area of science. On the other hand, the study of Barbalios et al. (2013) also took place in the above-mentioned area, specifically in environmental education. Its design consisted of using a realistic virtual environment with 3D technology where water simulations are developed in order to help students to achieve cognitive advances on ecosystems and acquire complex abstract notions with respect to other groups where other types of instruction were used. Unlike previous studies, the study conducted by Pareto (2014) focused on the area of mathematics with students from second to sixth grade. Using the game as the core on which the whole study is based, this research encourages an approach to discover knowledge in a playful way, providing the possibility of incorporating the teachable agents as an extension of the game in which students are guided and orientated so that they acquire learning through a set of questions, while giving them the possibility
of reflecting on mathematical learning.

Advancing from the educational stage, the studies of J. Anderson and Barnett (2011); Civelek et al. (2014); Fidan and Tuncel (2019); Ibáñez et al. (2014); Pellas and Vosinakis (2018); Riess and Mischo (2010); Shegog et al. (2012); Stieff (2011); Tatli and Ayas (2013) and Walker et al. (2014) were carried out in secondary education; and Olde et al. (2013), in vocational education. Regarding the first one, the study by) J. L. Anderson and Barnett (2013) used the Supercharged! application, based on the video game to teach complex concepts about electromagnetism. Among their main findings, they highlight the effectiveness of using such applications for the consolidation of complex concepts and the development of higher order metacognitive skills. Also related to the teaching of electromagnetism notions is the research of Ibáñez et al. (2014), which uses AR and whose findings were similar. In turn, in the area of physics and the use of simulations to achieve better academic results among students, the research by Civelek et al. (2014) with K-12 students should be emphasised. The contribution of this work, compared to others, is the measurement of student achievement, together with the attitude of students towards physics. In this regard, not only do simulations and, therefore, the AI, favour student learning, but there is also a change in student attitude towards STEM (Science, Technology, Engineering, and Mathematics) matters. In this line, it is also worth mentioning the study of Shegog et al. (2012) in the area of biology for the learning of transgenic animal models through simulations, finding that their use, apart from allowing access to certain concepts which otherwise would not be possible, reported improvements in achievement among students and advances in the students’ attitude towards the study of biology. In addition, the study carried out by Fidan and Tuncel (2019), which links AR to problem-based learning, explores the effect of these strategies on student achievement and their attitude towards physics, while providing teachers with some keys for incorporating these strategies into the instructional processes. Also contextualised in problem solving, but in this case, using simulation games, is the study conducted by Pellas and Vosinakis (2018). The research was implemented in a programming course with K-12 students, where it was determined that the use of OpenSim with Scratch4SL led to the development and improvement of computational strategies and more correct codes in problem solving. Within the contextualised studies in science, stands out the research of Stieff (2011), who used the simulations to facilitate the understanding of both molecules and representational skills in the area of chemistry. Specifically, the implementation of Connected Chemistry revealed better results in achievement compared to other modes of instruction, although the curriculum and the teacher were positioned as other decisive factors when interpreting these findings. In contrast, students in Connected Chemistry showed a tendency to use representations in the post-test. Developed in a virtual chemistry laboratory environment, the study by Tatli and Ayas (2013) found a similar effectiveness between the virtual laboratory and the real laboratory in terms of achievement. In the area of biology, through simulations for sustainable education, the research by Riess and Mischo (2010) revealed quantitative improvements in student achievement in the computer simulation condition, as well as in the combination of computer simulation and lesson condition with respect to the control group. As for vocational education, the study of Walker et al. (2014),
aimed at analysing the effects of intelligent peer tutoring on the quality of students’ collaborative and perceived support interactions and the effects they report on their learning, suggests a significative improvement regarding other non-adaptive modalities.

In Higher Education, there are studies such as Bortnik et al. (1968), which demonstrated the positive effects of implementing an “adopted approach” combining both virtual and practical learning environments. This had the potential to improve students’ research skills and practice in analytical chemistry studies. In this vein, Yelamarthi and Drake (2014) analysed the effect of introducing simulations in Digital Circuits Course with engineering students and found that students improved their performance and their interest towards learning thanks to the combination of active strategies and online preview of lectures, face-to-face student/instructor and peer interactions, discussions and hands-on activities. The study carried out by Jiménez-Hernández, Oktaba, Díaz-Barriga, and Piattini (2020) implemented an experiment to prove the effectiveness of web-based gamified software in the use of Booleans in a b-learning situation, with improvements in both performance and motivation in the experimental group.

Studies based on AI are also being developed in the field of medical education. For example, the study conducted by Veredas et al. (2014) compared the performance of students in the ePULab modality versus the traditional one and found better academic results in the experimental group.

For their part, Dickerson and Clark (2018) used the simulation tool SPICE in a micro-electronics course, finding improved performance and positive disposition towards active learning. Similarly, the study conducted by Neri et al. (2018) within a mechanics course involving visuo-haptic simulators reported better results in both achievement and motivation compared to the control groups. Using the potential of simulations in education, the study made by Fang and Guo (2016) contextualised in an undergraduate engineering dynamics course informed improved outcomes in terms of conceptual and procedural learning. Using AR, the research undertaken by Madathil et al. (2017) analysed the achievement, perceived commitment, user-friendliness and satisfaction of 165 university students enrolled in technical careers. They designed 2 experimental conditions: one based on VR; and the other, on a case study with photos and one control group. The experimental group performed better in terms of achievement than the other two groups and obtained higher scores in terms of commitment, ease of use and satisfaction than the other experimental group, although it matched the control group. The research by Zacharia and Olympiou (2011) confirmed the importance of physical and virtual manipulation in learning acquisition.

In this regard, J. Anderson and Barnett (2011), who used video games to teach the principles of electromagnetism to future teachers with Supercharged!, found different results to those obtained by the authors in secondary schools (J. L. Anderson & Barnett, 2013). In this case, even though the future teachers in the experimental group performed the tasks better, their scores were lower than those of the control group.

Finally, Bozkurt and Ilik (2010) found that computer simulations improved both student achievement and their attitude towards physics education.
3.2 Qualitative Meta-analysis

Figure 2 shows the distribution of effects of the included studies using a forest plot for the pre-post meta-analysis (panel A) and for the post meta-analysis (panel B). Of the 25 studies that were included in the analysis, 11 provided information to estimate the effect size for the pre-post meta-analysis. The average effect was $g_\Delta = 0.552$, 95% CI [-0.046, 1.150], not being statistically significant; $z = 1.809$, $p = .070$. The level of heterogeneity was significant: $Q(39) = 785.498$, $p < .001$. Regarding the standardized mean change difference scores for post-test meta-analysis, the mean effect was significant: $g_p = 0.716$, 95% CI [0.426, 1.007], $z = 4.833$, $p < .001$. And the effect distribution revealed significant heterogeneity: $Q(53) = 685.060$, $p < .001$.

Figure 3 represents the distribution of estimated effect sizes for the studies included in $g_\Delta$ (panel A) and the studies included in $g_p$ (panel B) using a funnel plot. As can be seen in the figure, both distributions show a high asymmetry in the distribution of effects. The red line represents Egger’s regression test. This test analyses the asymmetry that shows the distribution of the effects for each analysis. Both regressions showed to be significant: $b_1 = 15.662$, $z = 11.395$, $p < .001$ in the case of $g_\Delta$, and $b_1 = 18.397$, $z = 14.483$, $p < .001$ for $g_p$.

Finally, the moderating role that AI types could indicate on the total effect was analysed. Table 2 summarises the results of the moderator analysis for $g_\Delta$ and $g_p$. As it can be seen in the Q-test results, the AI type did not show a significant difference on the effect size for the analysis of $g_\Delta$; on the contrary, it exhibited a significant effect size for the analysis of $g_p$. Numerically, the type of AI VR tends to present an effect size ($g_\Delta = 2.01$ and $g_p = 1.28$) on its impact on learning that is higher than the other types of AI manipulation. The Application type also reveal to be a manipulation associated with an acceptable effect size ($g_\Delta = 0.39$ and $g_p = 0.92$), showing a significant effect in $g_p$ analysis.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Moderator</th>
<th>$g$</th>
<th>LL</th>
<th>UL</th>
<th>$z$</th>
<th>$p$</th>
<th>$k$</th>
<th>$Q$</th>
<th>df</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$g_\Delta$</td>
<td>Type of AI</td>
<td>0.39</td>
<td>-0.63</td>
<td>1.42</td>
<td>0.75</td>
<td>.450</td>
<td>5.29</td>
<td>4</td>
<td>.25</td>
<td></td>
</tr>
<tr>
<td>Ap</td>
<td>VR</td>
<td>2.01</td>
<td>-0.01</td>
<td>4.04</td>
<td>1.94</td>
<td>.051</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim</td>
<td>AR</td>
<td>0.47</td>
<td>-0.55</td>
<td>1.49</td>
<td>0.90</td>
<td>.368</td>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.25</td>
<td>-1.21</td>
<td>1.73</td>
<td>0.34</td>
<td>.731</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$g_p$</td>
<td>Type of AI</td>
<td>36.86</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ap</td>
<td>VR</td>
<td>0.92</td>
<td>0.48</td>
<td>1.36</td>
<td>4.09</td>
<td>&lt;.001</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim</td>
<td>AR</td>
<td>1.28</td>
<td>0.61</td>
<td>1.94</td>
<td>3.77</td>
<td>&lt;.001</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.49</td>
<td>0.08</td>
<td>0.90</td>
<td>2.35</td>
<td>.018</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.20</td>
<td>-0.52</td>
<td>0.93</td>
<td>0.54</td>
<td>.582</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note.** VR: Virtual Reality; AR: Augmented Reality; Ap: Application; Sim: Simulations; $g$: effect size; $LL$: lower limit of the 95% CI; $UL$: upper limit of the 95% CI; $z$: z-score associated with the g value in the same row; $p$: p-value associated with the z-score in the same row; $k$: number of effect sizes contributing to $g$ in the same row; $Q$: result of the Q-test for moderation; df: degrees of freedom of the Q-test for moderation; $p$: p-value of the Q-test for moderation.
Figure 2 Forest plot for $g_\Delta$ (panel A) and $g_p$ (panel B)
4 DISCUSSION AND CONCLUSIONS

The systematic review carried out aimed at analysing the impact of different AI components and computational sciences on student performance. After applying the inclusion and exclusion criteria mentioned above, the search resulted in 25 quasi-experimental studies with experimental and control groups, and/or hybrids, incorporating, in addition, qualitative strategies, which examined the positive effects of applying AI and computational sciences with respect to traditional methods.

Figure 3 Funnel plots for gΔ (panel A) and gp(panel B)
Specifically, it was intended to provide answers to four research questions, which will be answered below:

### 4.1 Does EAI Improve Student Performance?

All the studies found have shown the usefulness of employing EAI-based methods over more traditional methods. Among the pedagogical intentions and considerations that justify the use of EAI-based methods, the importance of placing students into the heart of their learning, providing them with opportunities to take an active role in that construction, stands out. The pedagogical possibilities offered by EAI are oriented to achieve significant learning in students, encouraging the visual component that has its different modalities, such as simulations, VR, AR, or applications such as games.

One of the great benefits of EAI is the flexibility to adapt educational programs to the rhythms and circumstances of each student (Deng, Benckendorff, & Gannaway, 2019; Vilkova & Shcheglova, 2020). Considering the advances in society and technology, teaching and guidance by teachers may often be done in a non-face-to-face modality. These modalities are the basis of distance education. This type of education arises due to the incompatibility between individuals’ studies and their work, or to situations of major importance, such as the global pandemic we are experiencing. Moreover, one of the main advantages of using technology in distance education is video tutorials (Wilkie & Liefeith, 2020). Students obtain personalised feedback on their teaching-learning process through them. On the other hand, teachers can record their classes or brief extracts of any content and students can watch it as often as necessary for their proper learning (Elliot, Gehret, Valadez, Carpenter, & Bryant, 2020; López-Rodríguez & Barac, 2019).

### 4.2 What Effects Does EAI Have on Students?

Based on our research findings, we have identified that the different EAI modalities not only affect the quantity of what students learn, but also lead to higher levels of motivation, which is demonstrated by a greater willingness to be involved in their learning. At the same time, it has been shown that most studies on EAI have been contextualised in STEM knowledge areas, which require higher levels of abstraction and greater complexity to achieve a proper understanding of knowledge. Depending on the level of education, different resources can be used so that the students are encouraged to manipulate them. In this regard, EAI not only helps to keep students focused while they are building something, but also encourages their creative ability to shape their thoughts (Barak & Zadok, 2009). Likewise, several studies use and have demonstrated the effectiveness of the AI tools in education (Fabregas et al., 2016; Jiménez et al., 2010). In addition, many educational institutions around the world are employing the STEM teaching methodology. This methodology is characterised using a series of new and up-to-date tools for the teaching of different school subjects. Moreover, it allows the design and development of a computational model based on learning and teaching conditions controlled on any subject with a high visual and multimedia content, which facilitates the acquisition and understanding of the contents through the ongoing interaction with the computer (Vlachopoulos & Makri, 2017).
In some of the studies included, besides checking the effectiveness of EAI on the quantity and quality of student learning, it has been shown how these modalities have led to changes in students’ attitudes towards these knowledge areas. In this regard, it may be determined that incorporating technologies, resources and EAI strategies into teaching methodologies is a significant advance on the road to achieving a meaningful and integrated student learning.

4.3 What Type of AI and Computational Science is the Most Common in the Educational Field?

Regarding the type of AI modality used, it was found that the different studies included address the potential of EAI on student performance through applications (n = 8), simulations (n = 8), VR (n = 4) and AR (n = 2). In relation to the applications, the literature suggests that their use in education is increasingly widespread at all levels (Wirjawan et al., 2020). Given their high impact on students, subject-based learning is contextualised and updated to the digital society. Applications are contextualised to the subject matter and educational level of the student. There are applications that allow the user to interact with their environment, use question and answers, find some element, orientate, watch instructional videos, create a portfolio, learn mathematics, play video games, and even intelligent tutoring. Likewise, research can be found that focuses on the use of some application carried out at any educational stage (Dunleavy et al., 2019; Hoplock, Lobchuk, & Lemoine, 2020; Petko et al., 2019). It is also important to note that apps tend to be used more frequently by teachers, due to their greater accessibility compared to other AI modalities, such as Big Data, which requires a higher level of literacy and involve other variables such as ethical issues (Gao, Li, & Liu, 2021). We can also find how there are studies in which different applications are used with people with autism (Law, Dutt, & Neihart, 2019) or in people with attention deficit hyperactivity disorder (Butt et al., 2020; Păsărelu, Andersson, & Dobrean, 2020). This implies that this type of AI is easier to implement in the design, adjustment and development of instructional processes for learners with special educational needs (Zhai et al., 2010).

Regarding the simulations, they provide a variety of scenarios in which students can learn by discovering applying what they have learned to progress at different levels, playing games or solving day-to-day problems (Masson & Rennie, 2006). Their easy access, potential for individualisation and low cost are some of the benefits that encourage its widespread use in education (Cabero-Almenara & Costas, 2016).

Among the findings obtained in these studies, there are common results and objectives: many of them sought to understand complex concepts in subjects classified as “difficult” and to strengthen and improve the students’ attitude towards the subject and, in some cases, to establish collaborative strategies among students. Likewise, the research that used the simulations as an AI, AR and VR modality was oriented to offer more real learning situations for the students, demonstrating an effectiveness similar to that of real laboratories. These experiences allow students to interact with an outside environment within the classroom (Lau & Lee, 2015). Furthermore, the use of VR in education enables students to interpret signs,
whether visual, auditory, or haptic, and to build their knowledge through their movement and their interaction with their own environment (Beck, 2019; Vesisenaho et al., 2019). In order to implement this technology into practice at school, teachers need to experiment with it, learn how to use it and then contextualise the content to the students’ own environment (Tondeur, Roblin, Van Braak, Voogt, and Prestridge, 2017). In higher education, university studies in medicine or nursing include the use of VR to perform operations or treatments (Baxter & Hainey, 2019; Bernardo, 2017).

4.4 Is EAI Effective at All Educational Stages?
The systematic review and meta-analysis have included studies from different educational stages. Thus, there are three studies in primary education, eleven contextualised studies in secondary education, one in vocational studies and ten studies in university. In all of them, the effectiveness of AI has been shown. However, the superiority of research in secondary education and university seems to point to a tendency to develop more studies in advanced educational stages, due, among other reasons, to the potential intellectual growth of students at these stages.

Similarly, small differences in terms of the type of AI and computational science have been observed in the different educational stages. According to the results obtained, there is a more restricted trend towards the use of AI and computer science, with applications and simulations becoming the most widely used in primary education. On the other hand, in secondary education and university, there is a tendency to use more types of AI and computational science, also due to the wide variety of subjects that can be studied and the tendency to encourage more computational thinking in students than in previous educational stages. However, this differs from Sun, Guo, and Hu (2021), who state that the use of gamification strategies to enhance computational thinking tends to be more effective at lower educational stages, making gamification a factor to be considered in this analysis.

4.5 Strengths and Limitations of the Study
In this regard, there is a change of trend in relation to the development of studies that use AI and promote computational thinking in students due to the pressure from international organisations to include computational thinking in the curricula at all educational stages. Thus, an increasing number of research studies is being found in the literature that uses AI and promotes computational thinking at different educational stages (McCormick & Hall, 2022; Merino-Armero, González-Calero, & Cozar-Gutierrez, 2022). This paper has a set of limitations that need to be emphasized. Firstly, the limitation related to the lack of studies that empirically demonstrate the relationship between AI and student achievement was considered. Another issue to take into account is determined by the databases themselves. Problems are often perceived in the registration of manuscripts, depending on the type of document or language, which may imply any possible bias in the elaboration of the systematic review. Related to this, the choice to use exclusively the WOS and Scopus databases and to ignore other publication sources, such as ERIC, PSYINFO, or the grey literature, may lead to bias. However, it was considered appropriate to focus on them because of its prestige and
since they are the databases that include the most manuscripts in the research area. In addition, this paper also has several strengths that should be noted. For example, this review has quantitatively assessed the effect of different EAI-based studies on student performance. In addition, studies from different countries have been collected, which provides an overview of this topic.

### 4.6 Contribution of This Paper

Experiences of AI in education can be found in the literature. Most of the works highlight the potential of AI to facilitate the understanding of complex and abstract knowledge. However, the negative points emphasize that AI requires specific software and the design of materials and technological resources. This particular paper provides scientific evidence of the real benefits of incorporating AI into educational processes, based on the meta-analysis conducted. Specifically, the size of the effect of the different AI modalities and computational sciences on the academic performance of students has been measured, and the benefits have been found to be real. Despite the high economic cost, this work gathers experiences of AI in different educational stages, which opens the way towards the implementation of teaching methodologies based on AI in both early childhood and higher education, with a view to decreasing learning difficulties in students and increasing their motivation towards learning. Similarly, it is important to note the impact on teachers themselves. While it is true that the use of AI and computer science in the field of education can provide significant support for teachers, who have seen their roles increase over the last few years. However, it also brings with it several challenges that need to be clarified. Firstly, there is digital literacy, where they must know how these modalities work, their viability and usefulness in their subject, as well as how to extrapolate it to the reality of their classroom in order to build meaningful learning situations for their students. Second, it must consider the sustainable objectives proposed by UNESCO (2021) both in the design of learning situations and in their implementation, in order to ensure that ethical assumptions are met and that human potential is enhanced rather than hindered (Flores-Vivar & García-Peñalvo, 2023; UNESCO, 2019). Accordingly, challenges that teachers must take on

<table>
<thead>
<tr>
<th>Dimensions and proposals</th>
<th>Challenges</th>
<th>Educational stage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessibility</td>
<td>Design and use of new educational resources.</td>
<td>All educational stages</td>
</tr>
<tr>
<td>Intelligent tutoring</td>
<td>Monitoring of intelligent system-generated responses to ensure that the ethical guidelines of SDG 4 are met.</td>
<td>University</td>
</tr>
<tr>
<td>Virtual facilitators</td>
<td>Design of new pedagogical models that include AI and computer science.</td>
<td>All educational stages</td>
</tr>
<tr>
<td>Intelligent content</td>
<td>Ethical and legal issues related to intellectual property.</td>
<td>Secondary Education and University</td>
</tr>
<tr>
<td>Teacher and AI collaboration</td>
<td>Digital literacy plans.</td>
<td>All educational stages</td>
</tr>
<tr>
<td>Content management and analytics</td>
<td>Ethical issues associated with data processing.</td>
<td>University</td>
</tr>
<tr>
<td>Out-of-class tutoring</td>
<td>Human-machine interaction paradigm.</td>
<td>Secondary Education and University</td>
</tr>
<tr>
<td>Learning management automation</td>
<td>Resource optimization.</td>
<td>All educational stages</td>
</tr>
</tbody>
</table>
in the inclusion of AI in the field of education have been identified (see Table 3).

4.7 Recommendations for Further Research

Based on the findings and limitations identified, the study should be extended to other databases, such as ERIC or PsycINFO, or the grey literature, which provide a worldwide view of current research on AI and students performance. During the search for studies carried out, it has been observed that the problem has grown exponentially over time.

5 AUTHORS’ CONTRIBUTION

Conceptualization: Inmaculada García-Martínez and José María Fernández-Batanero; Data curation: Samuel P. León; Formal analysis: Samuel P. León; Funding acquisition: José María Fernández-Batanero; Investigation: Inmaculada García-Martínez and José Fernández-Cerero; Methodology: Inmaculada García-Martínez and Samuel P. León; Software: Samuel P. León; Writing – original draft: Inmaculada García-Martínez, José Fernández-Cerero and Samuel P. León; Writing – review & editing: Inmaculada García-Martínez and José María Fernández-Batanero; Supervision: All authors have reviewed and approved this manuscript.

ACKNOWLEDGEMENTS

Funded by: Spanish State Research Agency, Spain
Funder Identifier: http://dx.doi.org/10.13039/501100011033
Award: PID2019-108230RB-I00

REFERENCES

García-Martínez, Inmaculada; et al. Analysing the Impact of Artificial Intelligence and Computational Sciences


doi.org/10.1109/TE.2009.2024689


Pareto, L. (2014). A teachable agent game engaging primary school children to learn arithmetic


