applied
sciences

MDPI

Review

Artificial Intelligence for Student Assessment: A Systematic Review

Víctor González-Calatayud [1], Paz Prendes-Espinosa [2,*] and Rosabel Roig-Vila [3]

1 Department of Statistics, Mathematics and Computer Sciences, Miguel Hernández University, 03202 Elche, Spain; victor.gonzalezc@umh.es
2 Department of Didactic and School Organization, University of Murcia, 30100 Murcia, Spain
3 Department of General Didactic and Specific Didactic, University of Alicante, 03690 Alicante, Spain; rosabel.roig@ua.es
* Correspondence: pazprend@um.es; Tel.: +34-868884063

Featured Application: The work provides insight into how AI is being applied to student assessment.

Abstract: Artificial Intelligence (AI) is being implemented in more and more fields, including education. The main uses of AI in education are related to tutoring and assessment. This paper analyzes the use of AI for student assessment based on a systematic review. For this purpose, a search was carried out in two databases: Scopus and Web of Science. A total of 454 papers were found and, after analyzing them according to the PRISMA Statement, a total of 22 papers were selected. It is clear from the studies analyzed that, in most of them, the pedagogy underlying the educational action is not reflected. Similarly, formative evaluation seems to be the main use of AI. Another of the main functionalities of AI in assessment is for the automatic grading of students. Several studies analyze the differences between the use of AI and its non-use. We discuss the results and conclude the need for teacher training and further research to understand the possibilities of AI in educational assessment, mainly in other educational levels than higher education. Moreover, it is necessary to increase the wealth of research which focuses on educational aspects more than technical development around AI.

Keywords: artificial intelligence; assessment; education; student feedback; technology enhanced learning; educational innovation

1. Introduction

Advances in digital technologies and computer sciences are leading us towards a technological society where progressively machines are designed and developed to meet human needs while simultaneously becoming smarter. There is a general consensus [1–5] that Artificial Intelligence (AI) will be one of the most valuable technologies for the coming years, in conjunction with others such as robotics, virtual reality, 3D printing or networks.

As Curugullo [6] remarks, there is not a universal definition of AI, but we can easily agree that it describes the integration of artificial (not a natural process, but one induced by machines) and intelligence (skills of learning, to extract concepts from data and to handle uncertainty in complex situations). Finally, the author concludes that AI is "an artifact able to acquire information on the surrounding environment and make sense of it, in order to act rationally and autonomously even in uncertain situations" (p. 3).

The comparison between human and artificial intelligence is really interesting [7,8]. These authors consider that defining AI as always linked to human intelligence is a mistake [8], so they understand this vision as anthropocentric. On the contrary, they use a non-anthropocentric concept and they regard AI as "the capacity to realize complex goals", so "The fact that humans possess general intelligence does not imply that new,

inorganic forms of general intelligence should comply with the criteria of human intelligence" (p. 2) [8]. The main differences between both types of intelligence are related to these factors: basic structure (biology versus digital systems); speed (humans being slower than computers); connectivity (human communication being slower and more intricate than AI systems); updatability and scalability (AI capacity for immediate improvement or to upscale); and energy consumption (the human brain being more efficient because it consumes less energy). One of their conclusions is: "No matter how intelligent autonomous AI agents become in certain respects, at least for the foreseeable future, they will remain unconscious machines" (p. 9).

AI is applied to different contexts, so we can see applications to architecture in smart buildings and smart cities [6], smart mobility [9], medicine [10–12], smart industry (the 4th industrial revolution, as some authors [13] remark) and also smart education or smart classrooms [14].

In relation to ethical uses of AI, UNESCO [1] is working on the elaboration of a global framework to guide uses and applications of AI in order to ensure the ethical use of these emerging technologies. We must consider the multiple advantages but, simultaneously, we must anticipate risks, malicious uses, and divides to guarantee human rights and dignity.

So, it is evident that not only is AI a relevant topic of research and innovation in education, but in general, for the future of our society. We are going to explore the real state of their applications in student assessment, because it is one of the fields with more relevant and innovative AI applications in education, as we are going to prove. This paper is organized in five sections. The next two sections are focused on the theoretical framework, so we introduce the educational approach to AI and the uses of AI for assessment. Next, we explain our method to develop the systematic review based on PRISMA model (Preferred Reporting Items for Systematic Reviews and Meta-Analysis). Finally, we present the main results to sustain the discussion, conclusions, and future trends. The main contributions of this paper include:

- Current revision of the state of the art.
- The most recent applications of AI to assess students.
- Based on our analysis, we suggest the main ways to improve education using AI and make some predictions on the future of this field of research.

2. Educational Applications of AI

As we have said above, one of the most relevant applications of AI is in the field of education. We refer not only to face-to-face education and smart learning environments, but also, and principally, in e-learning, making real the automatic and personalized learning processes based on adaptive learning, machine learning, ontologies, semantic technologies, natural language processing or deep learning. The origin of automatic learning processes goes back to, and is based on B.F. Skinner's teaching machine and programed learning [15], but we have observed a big evolution in them ever since.

In 2019, a lot of experts in AI attended the Beijing Conference and they agreed upon the "Beijing Consensus on AI and Education" [16]. In this document, they highlight the relevance of promoting AI in education, as in the Sustainable Development Goal number 4 set by UNESCO [16]. The main actions to promote are the following: to plan educational policies; to promote uses for education management, to empower teachers and learners; to promote values and skills for life and work; to offer lifelong learning opportunities and break the barriers in the digital world; to encourage equitable and inclusive uses, including gender equality; ethical and transparent uses; to research, evaluate and support innovation.

There are some interesting real experiences around AI, and the Horizon Report [17] compiles some of them, such as Watson Tutor by IBM, which has been tested with teachers and students in Texas or with Canvas LMS; another experience is from UC San Diego, the DSMLP (Data Science/Machine Learning Platform), which uses machine learning to provide access to resources and student projects; the third example is Boost, a mobile app integrated with Canvas LMS, a personal assistant for online learning; or finally, the

intelligent tool Edulai, for assessing and developing soft skills. These experts [17] consider that "despite ethical concerns, the higher education sector of AI applications related to teaching and learning is projected to grow significantly" (p. 27).

Another relevant report [16] analyzes three areas: learning with AI (the use of these tools in education), learning about AI (how to use these tools) and preparing for AI (understanding their potential to change our lives). In relation to the use of these tools in education, we can find diverse experiences, but overall, these are related to tutoring or assessment.

2.1. AI for Tutoring

We can find interesting studies about intelligent tutoring systems [13,18], which report that its development goes hand in hand with human natural language processing systems and learning analytics. All these systems are based on the idea of efficient mechanisms to provide feedback with enough quality to complement the teacher's action and substitute it in some cases. Considering the diversity of complex tasks that students can develop, it is a relevant advance to understand that these systems provide personalized answers. Ocaña et al. [13] remark that AI will mainly impact the process of personalized education because of the automated assistance, especially in the context of virtual interaction, and they include some very well-known examples such as Duolingo, which is the visible demonstration of the accessibility of applications based on the interaction between machines and humans, relying on AI principles.

Narciss et al. [19] remark that personalized tutoring feedback is one of the fields with more educational applicability, related to computer-based technologies. These authors studied a web-based intelligent learning environment used in mathematics to correct tasks and trace students' progress. They concluded that are some differences in feedback efficiency related to gender, because firstly, females take more advantage from tutoring feedback conditions (especially when feedback is based on conceptual hints). Secondly, girls improved their perceived competence more than boys, and finally, boys showed an increase in intrinsic motivation.

Jani et al. [20] demonstrate the usefulness of IA in both feedback and assessment as well as for formative evaluation, by using machine learning and checklists. Its research showed that automatic answers were good strategies to track students' progress and to identify areas to improve clinical practices.

The experience of Santos and Boticario [21] is of special interest because it utilizes AI to support collaborative learning experiences. They propose a Collaborative Logical Framework (CLF) based on AI to promote interaction, discussion, and collaboration as educational learning strategies. They also use this intelligent support to monitor students' behavior, reducing the workload in teachers. It is an adaptive guidance system developed for an e-learning platform (dotLRN) that can help the control and management of students' collaboration.

Ocaña-Fernández et al. [13] studied AI applications in higher education, and they conclude that intelligent tutors "have offered supportive assistance on several topics from their modest origins; topics such as training in geography, circuits, medical diagnosis, computing and programing, genetics and chemistry" (p. 563) and that they can be basic tools to improve the presumably ubiquitous online learning in the future. Some of the applications for providing feedback and tutoring are also thought for assessment, for instance Samarakou et al. [22] or Saplacan et al. [23].

2.2. AI for Educational Assessment

We can find interesting applications of AI linked to feedback. Feedback is a concept related to cybernetics, control, system theory, and it is in the origin of AI because it is the simplest way to understand the communication between humans and machines.

Mirchi et al. [11] used AI for simulation-based training in medicine, and they created a Virtual Operative Assistant to give automatic feedback to students based on metrics of

performance. From a formative educational paradigm, they integrate virtual reality and AI to classify students in relation to proficiency performance benchmarks and the system gives feedback to help them improve. In the same field of medicine, we can find the results of Janpla and Piriyasurawong [3]. They analyze the use of AI to produce tests in e-learning environments and they develop an intelligent software to select questions for online exams.

In a different work, Saplacan et al. [23] suggest that feedback provided by digital systems in learning situations have some problems such as eliciting negative emotions (these are neglect, frustration, uncertainty, need for confirmation and discomfort) experienced by students in higher education. Their research used a qualitative design based on a story dialogue method and its final conclusion is that "digital interfaces should also arouse positive emotions through their design" (p. 428).

The research of Samarakou et al. [22] is focused on continuous monitoring and assessment (Student Diagnosis, Assistance, Evaluation System based on Artificial Intelligence, StuDiAsE) of engineering students and AI proves its usefulness to provide personalized feedback and evaluate performance with quantitative and qualitative information.

Finally, the study of Rodríguez-Ascaso et al. [24] is focused on adaptive learning systems and self-assessment: "In the next future, we can expect that, within adaptive e-learning systems, both automatic and manual procedures will interoperate to elicit users' interaction needs for ensuring accessibility" (p. 1). They combine a personalized system with self-assessment and the use of learning objects for people with disabilities: "students with visual, auditory, and mobility impairments, as well as non-disabled students" and "results indicate that the procedure allows students, both disabled and non-disabled, to self-assess and report adequately their preferences to access electronic learning materials" (p. 8). They found, however, some interaction problems in a number of students with visual impairments.

2.3. Other Educational Uses of AI

We can also find other works related to education [4], such as data management, use or management of digital resources, experiences with SEN (Special Education Needs), predicting student achievement [25], or simply intelligent systems to answer questions. Chatterjee and Bhattacharjee [26] analyze AI to manage resources on educational organizations. Liu et al. [27] use this type of technologies to evaluate educational quality. Other authors [28–30] propose AI to personalize contents for students. We can develop intelligent systems to predict academic achievement [31]. In the same way, Ocaña-Fernández et al. [13] provide data on the possibilities of promoting applications that personalize education, adjusting to the individual needs detected by artificial intelligence algorithms, and thereby providing solutions, support and educational measures that respond to adaptive education models. These authors include some examples, such as chatbots (or bots), to interact in real time with users and provide personalized information.

Considering all this theoretical framework, this study focuses on how AI is being applied at a practical level for student assessment. Although mention is made of the AI tools used, the study aims to show the results of the research at a pedagogical level, so that anyone involved in education who wants to implement AI for assessment can do so on the basis of previous research.

3. Materials and Methods

This research is based on the use of a systematic review as a method to answer a research question through a systematic and replicable process [32]. Specifically, the PRISMA Statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) will be used as a model for this review [33]. Once the inclusion selection process has been carried out, based on pre-established criteria, the main results of the selected works are codified and extracted in order to synthesize and provide an answer to the question of how artificial intelligence is being used for student assessment.

3.1. Research Questions and Objectives

Our research problem is: what is the main use of AI in education? Around this problem, we can ask these research questions: Are there studies about student assessment based on AI applications? Where do the main studies come from? What is more relevant in this research around AI, the educational approach or the technical development? What type of student assessment is based on AI?

To answer these research questions, we focus on the next research objectives:

RO1. Identifying the main studies around student assessment based on AI in the last decade (2010–2020), using a systematic review.
RO2. Analyzing the impact that education and/or technology have on this field of research.
RO3. Analyzing the type of educational assessment which is being improved with AI.

3.2. Eligibility Criteria

This review includes research papers describing the use of Artificial Intelligence (AI) for student assessment, published in peer-reviewed journals, and published between 2010 and 2020. The year 2010 has been used as the starting year for the search, due to the great development of this technology derived from the beginning of the use of Siri by Apple. The languages used for the search were Spanish and English. Once the search had been carried out, the inclusion and exclusion criteria used were those set out in the following Table 1. It is important to clarify that the review is based primarily on the use of AI for student assessment in online and face-to-face subjects.

Table 1. Inclusion and exclusion criteria.

Inclusion Criteria	Exclusion Criteria
Published 2010–2020	Published before 2010
English or Spanish language	Not in English or Spanish
Empirical research	Not empirical (e.g., review)
Peer review journal	Not peer review journal
Use of artificial intelligence to assess learners	Not artificial intelligence
	Not learning setting
	Not for assessment

3.3. Information Sources

Studies were identified by searching electronic databases published in English and Spanish. Specifically, the search was conducted in the two main social science databases: Web of Sciences and Scopus. All relevant articles were accessed, with those that were not accessible needing to be excluded from the review.

3.4. Search Strategy

To ensure sensitivity in the search process, a co-occurrence analysis of terms was carried out to determine which terms were most common in relation to the research question [34]. From this analysis we concluded that the most common terms used for our search were: 'Artificial Intelligence', 'Machine Learning', 'Education', 'Assessment' and 'Evaluation'. To make the search with these terms more concrete, a protocol was designed with the combination of these terms and the use of Boolean operators. The combinations used in the two databases were:

Web of Sciences: (AB = (artificial intelligence OR machine learning) AND AB = Education AND AB = (assessment OR evaluation)). Type of document: articles. Time period: 2010–2020. Languages: English and Spanish.

Scopus: (TITLE-ABS-KEY (artificial AND intelligence) OR (TITLE-ABS-KEY (machine AND learning) AND TITLE-ABS-KEY (education) AND TITLE-ABS-KEY (assessment) OR TITLE-ABS-KEY (evaluation)). Type of document: articles. Time period: 2010–2020. Languages: English and Spanish.

The first search was conducted in March 2021 and found a total of 454 initial papers. After eliminating duplicates, a total of 449 articles were analyzed by the researchers.

3.5. Study Selection

The articles were evaluated and selected according to the above-mentioned criteria. Titles and abstracts were evaluated independently by two reviewers. Once selected, the papers were analyzed in full. When there was a discrepancy between the two reviewers, a third party assessed the proposal. To carry out the analysis of the studies, the Rayyan tool was used, which allows the review to be carried out using the "blind" mode.

3.6. Coding, Data Extraction and Analysis

For the coding of the articles, the same review tool Rayyan was used, which allows the establishment of "tags" for the categorization of the papers. For the extraction and analysis of the main information of the papers (based on the article data and the fundamental description of the use of AI for evaluation), a spreadsheet and the Mendeley manager were used.

4. Results

Of the 449 articles reviewed after eliminating duplicates, a total of 339 studies were excluded because they did not comply with the established criteria after reading the abstract and title. The full text of the remaining 100 papers was reviewed and a total of 68 were discarded for not meeting the criteria, and 10 for not having access to them. The workflow is shown in Figure 1 below.

The following Table 2 shows the main data of the selected studies: the authors, the year of publication, the journal, the authors' country of origin, their professional affiliation, and the subject or educational level at which AI was used are highlighted.
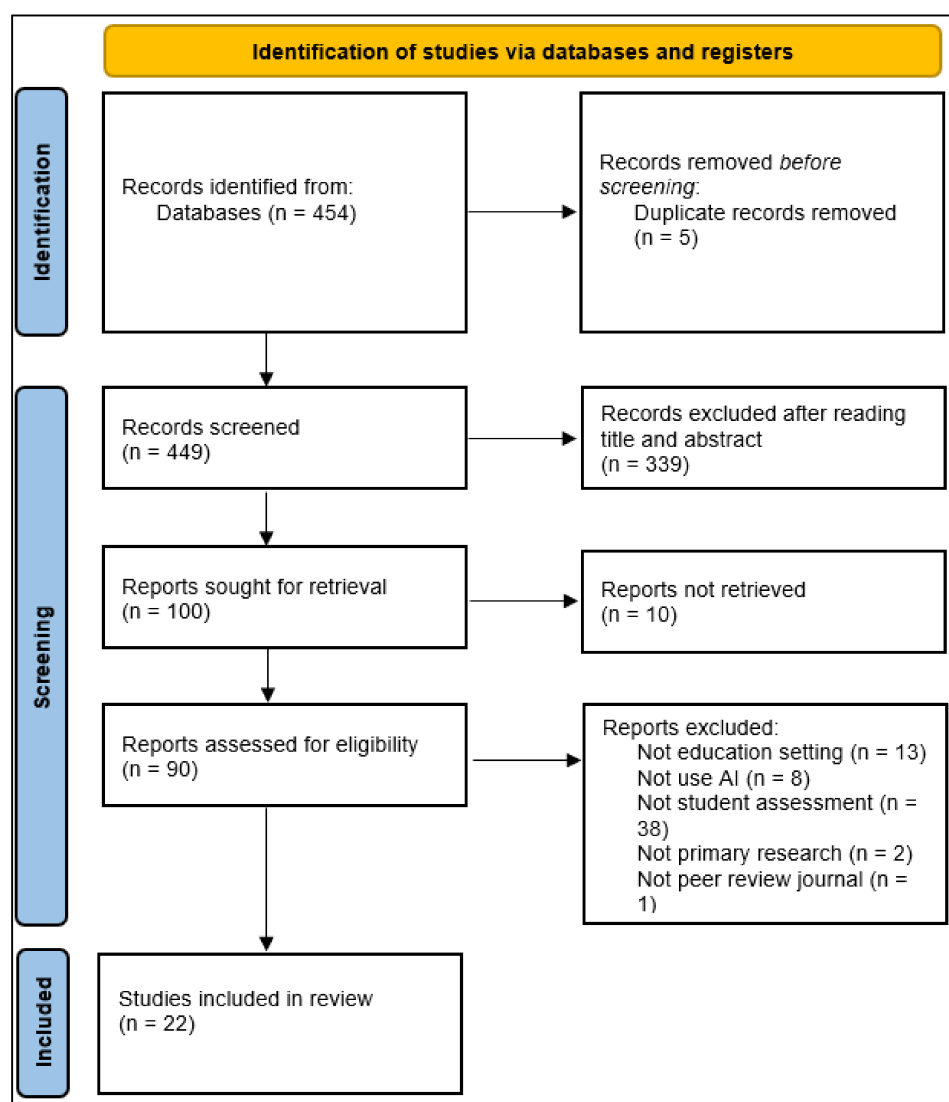
Figure 1. PRISMA diagram [31].

Table 2. Main data of the selected articles.

Authors	Year	Journal	Country	Author Affiliation	Subject or Educational Level
Rhienmora et al. [35]	2010	Methods of information in medicine	Thailand	Engineering and dentistry	Dental study
Phattanapon et al. [36]	2011	Artificial intelligence in medicine	Thailand	Computer science and Dentistry	Dental study
Gálvez et al. [18]	2013	International journal of artificial intelligence in education	Spain	Computer engineering	Computer science
Santos & Boticario [21]	2014	The scientific world journal	Spain	Computer science	Secondary education and University
Ouguengay et al. [37]	2015	Journal of theoretical and applied information technology	Morocco	Computer science	Amazigh Language

Table 2. *Cont.*

Authors	Year	Journal	Country	Author Affiliation	Subject or Educational Level
Samarakou et al. [22]	2016	Knowledge Management & E-Learning	Greece and UK	Engineering	Heat transfer
Kaila et al. [38]	2016	ACM Transactions on Computing Education	Finland	Information technology	Computer science, mathematics, and physics
Rapanta & Walton [39]	2016	International Journal of educational research	Portugal and Canada	Philosophy	Business and Education
Liu et al. [27]	2017	International Journal of Distance Education Technologies	China	Computer and information Science	English language
Perikos et al. [40]	2017	International journal of artificial intelligence in education	Greece	Computer engineering and informatics	Logic
Goel & Joyner [41]	2017	AI Magazine	USA	Interactive computing	Artificial intelligence
Grivokostopoulo et al. [42]	2017	International journal of artificial intelligence in education	Greece	Computer engineering	Artificial intelligence
Wiley et al. [43]	2017	International journal of artificial intelligence in education	USA	Psychology, Artificial intelligence	Secondary education, global warming
Malik et al. [44]	2017	EURASIA Journal of Mathematics Science and Technology Education	Pakistan	Information technology	English
Maicher et al. [45]	2019	Medical teacher	USA	Medicine	Medicine
Sun et al. [46]	2020	Computer intelligence	China and India	Engineering	English
Cruz-Jesus et al. [47]	2020	Heliyon	Portugal	Information management	Secondary education
Deo et al. [48]	2020	IEEE Access	Australia, Vietnam y Sweden	Engineering	Mathematics
Ince et al. [49]	2020	International Journal of Information Technology & Decision Making	Turkey	Vocational school of technical Science, Computer engineering	Mathematics
Ulum [50]	2020	Education and information technologies	Turkey	English language	English
Jani et al. [20]	2020	Medical education	USA	Medicine	Medicine
Choi & McClenen [51]	2020	Applied sciences	Korea and Canada	Adolescent coaching counselling, Computer science	Statistics

As can be seen in the table above, the journal with the most publications on this subject is the International Journal of Artificial Intelligence in Education, with up to four articles published in it. It is also striking that most of the authors are affiliated with areas of knowledge related to STEM subjects (Science, Technology, Engineering and Mathematics). However, the subjects in which AI was applied reveal, not only its use in those related to STEM, but also a wide use in language teaching, and in subjects related to health sciences.

The main use of AI for assessment is focused mainly on higher education contexts, although there are some examples of its use at secondary education level [43,47].

With regard to the articles in relation to their year of publication, there has been an increase in the number of publications in recent years, except in 2019 where only one article has been found. However, the year 2020 is the year in which the most papers are published in relation to this topic. Regarding the origin of the authors of the papers, it can be seen that the country which predominates in this type of study is the USA, followed by Greece. The following Table 3 lists all the authors of the selected papers.

Table 3. Number of contributions per country.

Country	Number of Contributions
USA	4
Greece	3
Thailand	2
Spain	2
Portugal	2
Chine	2
Canada	2
Turkey	2
Morocco	1
UK	1
Finland	1
Pakistan	1
Australia	1
Vietnam	1
Sweden	1
Korea	1

4.1. Understanding of AI

From the analysis of the various studies, it is clear that AI is a tool that can be applied in different ways and at different levels. It is, therefore, remarkable that only two studies begin the paper by stating their perspective on what AI is. Thus, the work by Sun et al. [46] indicates that "artificial intelligence (AI) is the combination of intelligence, that is, machines capable of demonstrating human intelligence and of making decisions with human skills"; AI is therefore intended to "build highly advanced machines that can make smart decisions" (p. 2).

Deo et al. [48] determine that AI could be incorporated as a subcategory of intelligent algorithms implemented in stand-alone models, or within a hybrid integrated computer systems model. Both studies agree on the idea that they are systems that prove to be efficient in making decisions similarly to what a human would do. The AI techniques used in the studies are diverse, although the most widely used is the fuzzy-logic system [22,37,49], often in combination with other methods.

4.2. Pedagogical Model Used

From the analysis of the different studies, it can be seen that the pedagogical aspect of AI application remains relatively underrepresented. There is only one study [18] that clearly addresses aspects related to pedagogical models that support the work to be developed subsequently. Specifically, it is based on three key concepts for the development of the platform using AI: constraint-based modeling, Items' response theory and Evidence Centered Design. Three more articles mention pedagogical models on which the instruction is based: two use collaboration as a method of student work [21,38] and the third uses argument maps [39].

Other works also touch on pedagogical aspects, although only descriptively in the theoretical framework. Thus, two of the works talk about this type of AI-based assessment for the development of competences in the subject [35,36], three of them determine the

need to give early feedback as an essential aspect in the educational process [22,37,42], one of the works determines that error detection is essential to achieve a better understanding and for pedagogical decision-making [40] and, finally, another work determines in the conclusions that it is necessary to incorporate a critical pedagogical stance at all levels for a better incorporation of this technology in education [50].

As can be seen from the studies highlighted in this section, most of those that mention pedagogical aspects are studies that are more than four years old. Recent work focuses more on the technical aspects of incorporating AI than on the pedagogical models that can underpin their use in education.

4.3. Formative Evaluation as the Reason for the Use of AI

Of the 22 selected papers, 15 have the use of AI for formative assessment as their starting point. Some of them explicitly mention this type of evaluation, while others can be extrapolated from what is described in the work. Three of the papers that do not use formative assessment focus on the use of AI in English language teaching [27,44,50], another two used AI in secondary education [43,47] and the other two in mathematics, but at university level [48,49].

Four studies specify that they carry out formative evaluation. In the case of Santos and Boticario [21], they are based on formative assessment through collaboration using a system called dotLRN. This system guides students in their collaborative work process. In the work of Goel and Joyner [41], it is clear that AI is a very powerful resource that allows frequent feedback to be given automatically to students and it can help teachers in this task when they have a large number of students. Maicher et al. [45] found that the use of AI and formative assessment can guide learners in the acquisition of competencies, in their case regarding the collection of medical information from the patient. Finally, with the work of Choi and McClenen [51], it is clear that for the personalization of teaching in e-learning, the incorporation of AI, in their case through the Computerized adaptive testing (CAT) method, it is a great resource to be considered.

4.4. Automated Scoring

In addition to the feedback that AI tools can provide, several studies have shown that one of the functionalities pursued is the automatic grading of students. In this way, we find up to ten studies that explicitly state that the system used grades students according to the task performed, although some of the other selected studies probably also use this tool, although it is not mentioned.

The work presented by Rhienmora et al. [36] is based on the assessment of competence in performing different dental techniques. The system used combines VR with AI. Based on the movements made by the user, the system evaluates his or her competence and determines a score, which leads to their categorization as either novice or expert. The study by Ouguengay et al. [37] assesses reading and writing skills in Amazigh by means of test assessment. The system establishes a score based on the answers given to the items. Kaila et al. [38], based on active learning methods, set up different types of collaboration-based activities that are automatically assessed. Goel and Joyner [41] in their AI-based course for teaching AI set up a system for students to automatically receive graders that allow them to quickly visualize their results and also provide feedback.

Grivokostopoulou et al. [42] use AI to evaluate students' performance in programing and automatically give them a grade on the task performed. Liu et al. [27] show how their employed AI system is able to automatically evaluate students' engineering essays in English. Jani et al. [20] use Machine Learning to evaluate OSCE (Observed Structured Clinical Exams) transcripts and automatically establish a score. The study by Maicher et al. [45] goes along the same lines by automatically assessing and grading medical students' collection of patient information. Ulum [50] describes the Turkish State University's application for assessing students' English language proficiency. This AI-based application establishes a grade automatically. Finally, Choi and McClenen [51] describe their AI system for forma-

tive assessment, which performs an automatic grading of learners to give feedback and to adapt the following tasks accordingly.

It is remarkable that most of the works that mention this automatic gradation of students are from the last years analyzed.

4.5. Comparison between IA Use and Non-Use

An essential aspect of the use of AI for assessment is its accuracy in learning compared to traditional methods. Thus, we found three studies that compare their results regarding this aspect. The work by Goel and Jyoner [41] makes a comparison between teaching AI from the platform they have created with this technology and another without the use of it. Their results show that students who used AI performed better than those who did not. On the other hand, the work of Grivokostopoulou et al. [42] made a comparison between the results obtained with AI and the teaching assessments made by hand to check the accuracy of this technology. In general, the results were very similar and correlated with each other. When analyzing the result in relation to four levels related to student grades—very low, low, medium, good, and excellent—it was observed that the results were quite similar, only finding higher results as excellent by the teacher than by the AI. The last study to mention the difference between those who used AI and those who did not is that of Samarakou et al. [22]. The subject in which this study was carried out was required the completion of a written assessment. From the results obtained, they determined that those who used AI scored 17% higher than those who did not use it.

5. Discussion and Conclusions

AI is today one of the main technologies being applied in all fields and at all levels. However, in education its use is not widespread, probably due to a lack of knowledge on the part of users [4], but it is set out to become one of the main tools to be used. As already mentioned, the main uses of AI applied to education are related to tutoring and assessment [40], although we also find other examples of use such as personalization [22,41,46] or quality assessment [20,37,39].

In this systematic review we have focused on analyzing the use of AI applied to student assessment based on the collection of articles published in the two main current databases: Web of Sciences and Scopus. To answer the first research objective, we can summarize that there are 454 works on AI and education in the last ten years, but after using our inclusion and exclusion criteria, we find out that only 22 studies are mainly focused on student assessment. The majority of them have been published in the last five years, and in the USA.

One of the main conclusions drawn from the analysis is that, despite the differences between human and artificial intelligence [7,8], this systematic research shows us the potential to facilitate and improve education, both face to face, blended or in virtual environments.

Despite this, and notwithstanding all the great technological progress that AI has made in recent years, it is clear from the studies analyzed that pedagogical models are not usually conceived in the studies carried out. The authors who describe their work focus more on analyzing the resource itself, explaining the algorithm and the platform used, but do not emphasize the pedagogical underpinning behind the use of certain activities. Only one study [18] makes clear the use of the pedagogical model behind the choice of activities that AI will then evaluate. Three other studies [21,38,39] also hint at their pedagogical rationale, which is mainly based on collaboration. This is closely related to the fact that practically all the authors involved are linked to STEM or health science degrees, disciplines in which pedagogy is not very much considered. So, our answer to the research objective 2 is that the approach to AI is mainly technical, the educational approach is a secondary element in the majority of the works.

Another of the main conclusions to be drawn from the analysis of the papers is the answer of our research objective three. We found that most of them focus on formative assessment, either implicitly or explicitly mentioning it. It is striking that three of

the papers that do not use AI for formative assessment are linked to the assessment of English [27,44,50]. As for the papers that focus on formative assessment, it is clear that the idea of using AI for this task lies in the help that this technology can provide to teachers when they have a large number of students [41], or based on the idea that learning improves the more immediate the feedback given to students [21].

In some cases, this formative assessment is carried out through the automatic marking of students' work. Almost half of the works analyzed clearly describe how the AI system used automatically assesses and grades students, and how this event is used for feedback. On other occasions, the AI system used only grades the students at the end of the task, as is the case in the work of Liu et al. [27]. It is also clear that most of the works that mention this automatic grading are from 2019 and 2020. The fact that virtual teaching has been developed worldwide during the 2020s may have led to a breakthrough in the development of this type of technology.

Finally, it is worth noting as a conclusion of the articles analyzed how in those that make a comparison between the use of AI in one group with another that does not use it, the former obtains better results. In the same way, Grivokostopoulou et al. [42] show how their AI system is suitable for student evaluation and grading, as it is similar to teacher evaluation.

Despite all these advances, and what AI can do in the field of education, as Ulum's work [50] shows, this technology needs to be humanized. Research so far shows that a machine cannot assume the role of a teacher, and the way artificial intelligence works and carries out processes in the context of teaching is far from human intelligence [7] and partly due to the lack of transparency in decision-making algorithms [4]. In addition to this problem, we are faced with the difficulty of putting this technology into practice by all teachers. Several authors [8,16,52], have determined that AI requires specific training of our students as future professionals, because they need to understand the characteristics, possibilities and limitations of these intelligent systems. In this sense, Ocaña et al. [13] include diverse skills such as the following: computational thinking, programing, computing skills, and finally information and audiovisual skills. Thus, computational thinking is one of the main skills to be developed from early childhood education to prepare students for a world where programing, robotics and AI will be essential to develop their abilities [53]. The experiences that are being carried out in early childhood education for the development of computational thinking are essential [54–56]. Likewise, it is crucial to train teachers in the use of this technology [57], but not only on the basis of learning the tools, but also based on pedagogical reference models that give meaning to the development of classes. If we want to promote technology enhanced learning and smart environments using AI, teachers are an essential element of successful.

Certainly, our study has found a small number of articles focused on student assessment in relation to the big number of AI applications in last years, but we have applied the method rigorously to guarantee the validity of the systematic review in relation to our eligibility criteria (see Table 1 above). Generally speaking, the possibilities that AI offers to education are enormous, especially for tutoring, assessment and personalization of education, and most of these are yet to be discovered [58]. We must promote the collaboration [59] between education experts and AI experts, because we cannot understand the educational potential of technologies if we do not understand the educational context, the characteristics of educational interaction and the real users' needs. Technology and pedagogy must walk together if we want to understand the future of the advanced technologies in education [60], in order to understand the new education that will arrive in the next years.

References

1. UNESCO. Elaboration of a Recommendation on the Ethics of Artificial Intelligence. Available online: https://en.unesco.org/artificial-intelligence/ethics (accessed on 22 May 2021).
2. United Nations Transforming Our World: The 2030 Agenda for Sustainable Development/Department of Economic and Social Affairs. Available online: https://sdgs.un.org/2030agenda (accessed on 22 May 2021).
3. Janpla, S.; Piriyasurawong, P. The Development of an Intelligent Multilevel Item Bank Model for the National Evaluation of Undergraduates. *Univers. J. Educ. Res.* **2020**, *8*, 4163–4172. [CrossRef]
4. Prendes Espinosa, M.P.; Cerdán Cartagena, F. Tecnologías avanzadas para afrontar el reto de la innovación educativa. *RIED. Rev. Iberoam. Educ. Distancia* **2020**, *24*, 35. [CrossRef]
5. Chai, C.S.; Wang, X.; Xu, C. An extended theory of planned behavior for the modelling of chinese secondary school students' intention to learn artificial intelligence. *Mathematics* **2020**, *8*, 2089. [CrossRef]
6. Cugurullo, F. Urban Artificial Intelligence: From Automation to Autonomy in the Smart City. *Front. Sustain. Cities* **2020**, *2*. [CrossRef]
7. Cope, B.; Kalantzis, M.; Searsmith, D. Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies. *Educ. Philos. Theory* **2020**, 1–17. [CrossRef]
8. Korteling, J.E.; van de Boer-Visschedijk, G.C.; Blankendaal, R.A.M.; Boonekamp, R.C.; Eikelboom, A.R. Human-versus Artificial Intelligence. *Front. Artif. Intell.* **2021**, *4*. [CrossRef]
9. Paiva, S.; Ahad, M.; Tripathi, G.; Feroz, N.; Casalino, G. Enabling Technologies for Urban Smart Mobility: Recent Trends, Opportunities and Challenges. *Sensors* **2021**, *21*, 2143. [CrossRef]
10. Hwang, S.; Song, Y.; Kim, J. Evaluation of AI-Assisted Telemedicine Service Using a Mobile Pet Application. *Appl. Sci.* **2021**, *11*, 2707. [CrossRef]
11. Mirchi, N.; Bissonnette, V.; Yilmaz, R.; Ledwos, N.; Winkler-Schwartz, A.; Del Maestro, R.F. The Virtual Operative Assistant: An explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS ONE* **2020**, *15*, e0229596. [CrossRef]
12. Houwink, E.J.F.; Kasteleyn, M.J.; Alpay, L.; Pearce, C.; Butler-Henderson, K.; Meijer, E.; van Kampen, S.; Versluis, A.; Bonten, T.N.; van Dalfsen, J.H.; et al. SERIES: eHealth in primary care. Part 3: eHealth education in primary care. *Eur. J. Gen. Pract.* **2020**, *26*, 108–118. [CrossRef]
13. Ocaña-Fernández, Y.; Valenzuela-Fernández, L.A.; Garro-Aburto, L.L. Inteligencia artificial y sus implicaciones en la educación superior. *Propósitos y Represent.* **2019**, *7*. [CrossRef]
14. García-Tudela, P.A.; Prendes-Espinosa, M.P.; Solano-Fernández, I.M. Smart learning environments and ergonomics: An approach to the state of the question. *J. New Approaches Educ. Res.* **2020**, *9*, 245–258. [CrossRef]
15. Fry, E. Teaching Machine Dichotomy: Skinner vs. Pressey. *Psychol. Rep.* **1960**, *6*, 11–14. [CrossRef]
16. UNESCO. *Beijing Consensus on Artificial Intelligence and Education*; UNESCO: Paris, France, 2019.
17. Chew, E.; Chua, X.N. Robotic Chinese language tutor: Personalising progress assessment and feedback or taking over your job? *Horizont* **2020**, *28*, 113–124. [CrossRef]
18. Gálvez, J.; Conejo, R.; Guzmán, E. Statistical Techniques to Explore the Quality of Constraints in Constraint-Based Modeling Environments. *Int. J. Artif. Intell. Educ.* **2013**, *23*, 22–49. [CrossRef]
19. Narciss, S.; Sosnovsky, S.; Schnaubert, L.; Andrès, E.; Eichelmann, A.; Goguadze, G.; Melis, E. Exploring feedback and student characteristics relevant for personalizing feedback strategies. *Comput. Educ.* **2014**, *71*, 56–76. [CrossRef]
20. Jani, K.H.; Jones, K.A.; Jones, G.W.; Amiel, J.; Barron, B.; Elhadad, N. Machine learning to extract communication and history-taking skills in OSCE transcripts. *Med. Educ.* **2020**, *54*, 1159–1170. [CrossRef]
21. Santos, O.C.; Boticario, J.G. Involving Users to Improve the Collaborative Logical Framework. *Sci. World J.* **2014**, *2014*, 1–15. [CrossRef]
22. Samarakou, M.; Fylladitakis, E.D.; Karolidis, D.; Früh, W.-G.; Hatziapostolou, A.; Athinaios, S.S.; Grigoriadou, M. Evaluation of an intelligent open learning system for engineering education. *Knowl. Manag. E-Learning Int. J.* **2016**, *8*, 496–513. [CrossRef]
23. Saplacan, D.; Herstad, J.; Pajalic, Z. Feedback from digital systems used in higher education: An inquiry into triggered emotions two universal design oriented solutions for a better user experience. In *Transforming Our World through Design, Diversity and Education: Proceedings of Universal Design and Higher Education in Transformation Congress 2018*; IOS Press: Amsterdam, The Netherlands, 2018; Volume 256, pp. 421–430.
24. Rodriguez-Ascaso, A.; Boticario, J.G.; Finat, C.; Petrie, H. Setting accessibility preferences about learning objects within adaptive elearning systems: User experience and organizational aspects. *Expert Syst.* **2017**, *34*, e12187. [CrossRef]
25. Qu, S.; Li, K.; Wu, B.; Zhang, S.; Wang, Y. Predicting student achievement based on temporal learning behavior in MOOCs. *Appl. Sci.* **2019**, *9*, 5539. [CrossRef]

26. Chatterjee, S.; Bhattacharjee, K.K. Adoption of artificial intelligence in higher education: A quantitative analysis using structural equation modelling. *Educ. Inf. Technol.* **2020**, *25*, 3443–3463. [CrossRef]

27. Liu, M.; Wang, Y.; Xu, W.; Liu, L. Automated Scoring of Chinese Engineering Students' English Essays. *Int. J. Distance Educ. Technol.* **2017**, *15*, 52–68. [CrossRef]

28. Kim, W.-H.; Kim, J.-H. Individualized AI Tutor Based on Developmental Learning Networks. *IEEE Access* **2020**, *8*, 27927–27937. [CrossRef]

29. Villegas-Ch, W.; Arias-Navarrete, A.; Palacios-Pacheco, X. Proposal of an Architecture for the Integration of a Chatbot with Artificial Intelligence in a Smart Campus for the Improvement of Learning. *Sustainability* **2020**, *12*, 1500. [CrossRef]

30. Xiao, M.; Yi, H. Building an efficient artificial intelligence model for personalized training in colleges and universities. *Comput. Appl. Eng. Educ.* **2021**, *29*, 350–358. [CrossRef]

31. Castrillón, O.D.; Sarache, W.; Ruiz-Herrera, S. Predicción del rendimiento académico por medio de técnicas de inteligencia artificial. *Form. Univ.* **2020**, *13*, 93–102. [CrossRef]

32. Gough, D.; Oliver, S.; Thomas, J. *An Introduction to Systematic Reviews*; SAGE: Los Angeles, LA, USA, 2017.

33. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *71*. [CrossRef]

34. Galvez, C. Co-word analysis applied to highly cited papers in Library and Information Science (2007–2017). *Transinformacao* **2018**, *30*, 277–286. [CrossRef]

35. Rhienmora, P.; Haddawy, P.; Khanal, P.; Suebnukarn, S.; Dailey, M.N. A virtual reality simulator for teaching and evaluating dental procedures. *Methods Inf. Med.* **2010**, *49*, 396–405. [CrossRef]

36. Rhienmora, P.; Haddawy, P.; Suebnukarn, S.; Dailey, M.N. Intelligent dental training simulator with objective skill assessment and feedback. *Artif. Intell. Med.* **2011**, *52*, 115–121. [CrossRef] [PubMed]

37. Ouguengay, Y.A.; El Faddouli, N.-E.; Bennani, S. A neuro-fuzzy inference system for the evaluation of reading/writing competencies acquisition in an e-learning environnement. *J. Theor. Appl. Inf. Technol.* **2015**, *81*, 600–608.

38. Kaila, E.; Kurvinen, E.; Lokkila, E.; Laakso, M.-J. Redesigning an Object-Oriented Programming Course. *ACM Trans. Comput. Educ.* **2016**, *16*, 1–21. [CrossRef]

39. Rapanta, C.; Walton, D. The Use of Argument Maps as an Assessment Tool in Higher Education. *Int. J. Educ. Res.* **2016**, *79*, 211–221. [CrossRef]

40. Perikos, I.; Grivokostopoulou, F.; Hatzilygeroudis, I. Assistance and Feedback Mechanism in an Intelligent Tutoring System for Teaching Conversion of Natural Language into Logic. *Int. J. Artif. Intell. Educ.* **2017**, *27*, 475–514. [CrossRef]

41. Goel, A.K.; Joyner, D.A. Using AI to Teach AI: Lessons from an Online AI Class. *AI Mag.* **2017**, *38*, 48–59. [CrossRef]

42. Grivokostopoulou, F.; Perikos, I.; Hatzilygeroudis, I. An Educational System for Learning Search Algorithms and Automatically Assessing Student Performance. *Int. J. Artif. Intell. Educ.* **2017**, *27*, 207–240. [CrossRef]

43. Wiley, J.; Hastings, P.; Blaum, D.; Jaeger, A.J.; Hughes, S.; Wallace, P.; Griffin, T.D.; Britt, M.A. Different Approaches to Assessing the Quality of Explanations Following a Multiple-Document Inquiry Activity in Science. *Int. J. Artif. Intell. Educ.* **2017**, *27*, 758–790. [CrossRef]

44. Malik, K.R.; Mir, R.R.; Farhan, M.; Rafiq, T.; Aslam, M. Student Query Trend Assessment with Semantical Annotation and Artificial Intelligent Multi-Agents. *EURASIA J. Math. Sci. Technol. Educ.* **2017**, *13*. [CrossRef]

45. Maicher, K.R.; Zimmerman, L.; Wilcox, B.; Liston, B.; Cronau, H.; Macerollo, A.; Jin, L.; Jaffe, E.; White, M.; Fosler-Lussier, E.; et al. Using virtual standardized patients to accurately assess information gathering skills in medical students. *Med. Teach.* **2019**, *41*, 1053–1059. [CrossRef]

46. Sun, Z.; Anbarasan, M.; Praveen Kumar, D. Design of online intelligent English teaching platform based on artificial intelligence techniques. *Comput. Intell.* **2020**, 12351. [CrossRef]

47. Cruz-Jesus, F.; Castelli, M.; Oliveira, T.; Mendes, R.; Nunes, C.; Sa-Velho, M.; Rosa-Louro, A. Using artificial intelligence methods to assess academic achievement in public high schools of a European Union country. *Heliyon* **2020**, *6*, e04081. [CrossRef]

48. Deo, R.C.; Yaseen, Z.M.; Al-Ansari, N.; Nguyen-Huy, T.; Langlands, T.A.M.; Galligan, L. Modern Artificial Intelligence Model Development for Undergraduate Student Performance Prediction: An Investigation on Engineering Mathematics Courses. *IEEE Access* **2020**, *8*, 136697–136724. [CrossRef]

49. İnce, M.; Yiğit, T.; Hakan Işık, A. A Novel Hybrid Fuzzy AHP-GA Method for Test Sheet Question Selection. *Int. J. Inf. Technol. Decis. Mak.* **2020**, *19*, 629–647. [CrossRef]

50. Ulum, Ö.G. A critical deconstruction of computer-based test application in Turkish State University. *Educ. Inf. Technol.* **2020**, *25*, 4883–4896. [CrossRef]

51. Choi, Y.; McClenen, C. Development of adaptive formative assessment system using computerized adaptive testing and dynamic bayesian networks. *Appl. Sci.* **2020**, *10*, 8196. [CrossRef]

52. Collazos, C.A.; Gutiérrez, F.L.; Gallardo, J.; Ortega, M.; Fardoun, H.M.; Molina, A.I. Descriptive theory of awareness for groupware development. *J. Ambient Intell. Humaniz. Comput.* **2019**, *10*, 4789–4818. [CrossRef]

53. Roig-Vila, R.; Moreno-Isac, V. El pensamiento computacional en Educación. Análisis bibliométrico y temático. *Rev. Educ. Distancia* **2020**, *20*. [CrossRef]

54. Álvarez-Herrero, J.-F. Diseño y validación de un instrumento para la taxonomía de los robots de suelo en Educación Infantil. *Pixel-Bit Rev. Medios Educ.* **2020**, *60*, 59–76. [CrossRef]

55. González González, C.S. Estrategias para la enseñanza del pensamiento computacional y uso efectivo de tecnologías en educación infantil: Una propuesta inclusiva. *Rev. Interuniv. Investig. Tecnol. Educ.* **2019**. [CrossRef]

56. Recio Caride, S. Experiencias robóticas en infantil. *Rev. Interuniv. Investig. Tecnol. Educ.* **2019**, *12*. [CrossRef]

57. Cabero-Almenara, J.; Romero-Tena, R.; Palacios-Rodríguez, A. Evaluation of Teacher Digital Competence Frameworks Through Expert Judgement: The Use of the Expert Competence Coefficient. *J. New Approaches Educ. Res.* **2020**, *9*, 275. [CrossRef]

58. Prendes Espinosa, M.P.; González-Calatayud, V. Interactive environments for involvement and motivation for learning. In *Video Games for Teachers: From Research to Action*; Payá, A., Mengual-Ándres, S., Eds.; Mcragw Hill: Madrid, Spain, 2019; pp. 17–38.

59. Fernández-Díaz, E.; Gutiérrez Esteban, P.; Fernández Olaskoaga, L. University-School Scenarios and Voices from Classrooms. Rethinking Collaboration within the Framework of an Interuniversity Project. *J. New Approaches Educ. Res.* **2019**, *8*, 79. [CrossRef]

60. Burkle, M.; Cobo, C. Redefining Knowledge in the Digital Age. *J. New Approaches Educ. Res.* **2018**, *7*, 79–80. [CrossRef]