Received: 26 May 2023 | Accepted: 17 November 2023

DOI: 10.1111/bjet.13413

REVIEW

British Journal of Educational Technology BERA

Empathic pedagogical conversational agents: A systematic literature review

Elvis Ortega-Ochoa[1] | Marta Arguedas[2] | Thanasis Daradoumis[2,3]

[1]Doctoral School, Universitat Oberta de Catalunya, Barcelona, Spain

[2]Computer Science, Multimedia, and Telecommunication Faculty, Universitat Oberta de Catalunya, Barcelona, Spain

[3]Department of Cultural Technology and Communication, University of the Aegean, Mytilene, Greece

Correspondence
Elvis Ortega-Ochoa, Doctoral School, Universitat Oberta de Catalunya at Rambla del Poblenou, 154, Barcelona 08018, Spain.
Email: egortega@uoc.edu

Funding information
Universitat Oberta de Catalunya

Abstract

Artificial intelligence (AI) and natural language processing technologies have fuelled the growth of Pedagogical Conversational Agents (PCAs) with empathic conversational capabilities. However, no systematic literature review has explored the intersection between conversational agents, education and emotion. Therefore, this study aimed to outline the key aspects of designing, implementing and evaluating these agents. The data sources were empirical studies, including peer-reviewed conference papers and journal articles, and the most recent publications, from the ACM Digital Library, IEEE Xplore, ProQuest, ScienceDirect, Scopus, SpringerLink, Taylor & Francis Online, Web of Science and Wiley Online Library. The remaining papers underwent a rigorous quality assessment. A filter study meeting the objective was based on keywords. Comparative analysis and synthesis of results were used to handle data and combine study outcomes. Out of 1162 search results, 13 studies were selected. The results indicate that agents promote dialogic learning, proficiency in knowledge domains, personalized feedback and empathic abilities as essential design principles. Most implementations employ a quantitative approach, and two variables are used for evaluation. Feedback types play a vital role in achieving positive results in learning performance and student perceptions. The main limitations and gaps are the time range for literature selection, the level of integration of the empathic field and the lack of a detailed development stage report. Moreover, future directions are the ethical implications of agents operating beyond scheduled learning times and the adoption of Responsible AI principles. In conclusion, this review provides a comprehensive framework of empathic PCAs, mostly in their evaluation. The systematic review registration number is osf.io/3xk6a.

KEYWORDS

affective feedback, conversational agents, design principles, emotion, learning outcomes, learning performance, student perceptions, systematic literature review

INTRODUCTION

The chatbot industry has been on an exponential growth trajectory, with a valuation of 190.8 million USD in 2016, which is projected to reach a staggering 1.25 billion USD by 2025 (Research and Markets, 2022). The inception of conversational agents dates to over half a century ago, with the development of ELIZA by Weizenbaum (1966). Today, the buzz around the singularity in artificial intelligence (AI) has become more prevalent, especially after the publication of ChatGPT (OpenAI, 2022). Natural language processing (NLP) technology enables conversational agents to facilitate human–computer interaction through natural language.

In the e-learning field, it is imperative to recognize students' emotional states as they progress in their learning journey. Each student is unique, and there is no one-size-fits-all approach to education. Emotions are pivotal in shaping the interaction process and therefore must be taken into consideration when designing any methodological strategy or learning tool. Pedagogical agents, or autonomous characters, are used to create rich, face-to-face learning interactions in cohabiting environments with students (Johnson et al., 2000). These agents, also known as educational chatbots, can function as independent tools or be integrated into Intelligent Tutoring Systems (ITSs). They interact with students through various modes of communication including text, speech, graphics, haptics and gestures. Empathic Pedagogical Conversational Agents (PCAs) have emerged as an important approach for enhancing and personalizing the learning experience. Thus, this Systematic Literature Review (SLR) focused on empathic PCAs.

Rationale

Recent studies have shown that empathic chatbots can improve students' learning efficiency by providing instant feedback (Wu et al., 2020). However, there are numerous challenges in interaction, such as the need for chatbots to possess human-like interpersonal qualities and the risk of conversation breakdown due to poor conversation skills. To ensure effective interactions, conversational agents must be syntactically correct, empathic and knowledgeable, and should be context-aware (Prendinger & Ishizuka, 2005). Components like affect, emotion, tone and sentiment are used to create empathic agents that can evoke empathic reactions in users (Kusal et al., 2022). An empathic agent is “a synthetic character that evokes an

Practitioner notes

What is already known about this topic

- Emotions play a pivotal role in shaping the interaction process, making it essential to consider them when designing methodological strategies or learning tools.
- Empathic Pedagogical Conversational Agents (PCAs) have emerged as a crucial approach for enhancing and personalizing the learning experience (24/7) for pupils and supporting human teachers in their teaching process.
- Despite the creation of numerous empathic PCAs, there is a scarcity of Systematic Literature Reviews (SLRs) on their application in the educational field, particularly concerning the integration of emotional abilities in combination with the competencies of each subject.

What this paper adds

- It offers new insights into the design principles underlying the integration of the empathic field.
- It reviews different approaches for incorporating students' prior knowledge in real time.
- It provides a comprehensive and up-to-date overview of the research designs used for implementation, including quantitative, qualitative and mixed methods.
- It examines the factors that influence the effectiveness of empathic PCA in teaching and learning.
- It evaluates the types of feedback that enhance the impact of the empathic field on learning outcomes.

Implications for practice and/or policy

- It is crucial to grasp the topics that this paper introduces in order to effectively integrate new learning tools into any context.
- Techno-pedagogical designers seeking to gain insights into empathic PCAs will find immense value in this SLR, as it comprehensively covers each stage of the process.
- For future research endeavours, this study offers a wealth of ideas to draw upon, enabling researchers to address the challenges outlined and explore new avenues of investigation.

empathic reaction in the user" (Hall & Woods, 2006, p. 310). In this regard, empathic PCAs are agents designed to integrate emotional abilities and interact with students, and recent advancements in AI and NLP technologies have driven the development of agents with advanced conversational capabilities (Liu et al., 2022; Terzidou et al., 2018; Wu et al., 2020).

Although many empathic PCAs have been created, there is a dearth of SLRs regarding their application in the educational field, especially in their integration of emotional abilities. While there are many general SLRs on conversational agents, their findings serve as the foundation for more in-depth studies. Recent SLRs focused independently on the education and emotion field are:

- *Education*. Kuhail et al. (2022) analyses the educational field, platform, design principles, the role of chatbots, interaction styles, evidence and limitations. Huang et al. (2021) analyses the possible pedagogical, technological and social affordances enabled by chatbots in language learning. Okonkwo and Ade-Ibijola (2021) analyses the profile for chatbot

applications in the education domain, the benefits and challenges of implementing chatbots in an educational setting, and potential future areas of education that could benefit from using chatbots.
- *Emotion*. Kusal et al. (2022) provides a review of AI-based conversational agents; the authors discuss how conversational agents can simulate human behaviour by adding emotions, sentiments and affect to the context. Bilquise et al. (2022) focuses on the empathic chatbot development stage, mentioning that 1% of such agents are applied in education. Rapp et al. (2021) focuses on the human–computer interaction perspective of chatbot usage by investigating human-like chatbots' usability and user acceptance.

Previous review studies have made significant contributions to the literature, albeit their main emphasis was the educational field, platform, design principles, the role of chatbots, interaction styles, evidence and limitations (Kuhail et al., 2022); the pedagogical, technological and social affordances of chatbots (Huang et al., 2021); and determining the benefits and challenges of implementing educational chatbots (Okonkwo & Ade-Ibijola, 2021). Nonetheless, none of these studies have explored the crucial field of empathic PCA. Given the extensive research on PCAs, it is essential to conduct an SLR that sheds light on key aspects of empathic PCA, such as design principles, approaches for integrating students' prior knowledge (Tegos & Demetriadis, 2017), research methods, variables for evaluation and feedback types.

Objective

Considering the above fields and the research gaps, this study analysed contributions in the intersection of conversational agent-education-emotion and aimed to answer the following research questions (RQs):

- RQ1: What are the design principles for building an empathic PCA?
- RQ2: What empathic PCA set-ups integrate students' previous knowledge?
- RQ3: What research approaches are used in the implementation of empathic PCA?
- RQ4: What are the variables for assessing an empathic PCA?
- RQ5: What types of feedback from empathic PCAs impact learning outcomes?

The main goal was to describe the empathic PCA's design, implementation and evaluation stages. The design stage centred around principles and methods for incorporating students' prior knowledge. The implementation, on the other hand, delved into research methodologies. Finally, the evaluation stage concerned with assessing variables and feedback methods that influence learning results. For that, the populations were the reports that link the three components: conversational agent, education and emotion. The intervention was the search context; thus, some keywords were identified. The results showed a comparison of the reports regarding the research questions. The outcomes were agents' design principles, set-ups for integrating the students' previous knowledge, research approaches, assessment variables and feedback types that impact learning outcomes. Finally, all types of research approaches were useful, whether quantitative, qualitative or mixed.

This study contributes to five key areas by systematically analysing papers on empathic PCAs. First, it offers new insights into the design principles that underlie the integration of the empathic field. Second, it reviews the different approaches for incorporating the students' prior knowledge in real time. Third, it provides a comprehensive and up-to-date overview of the research designs used for implementation, including quantitative, qualitative and mixed methods. Fourth, it examines the factors that influence the effectiveness of empathic

PCA in teaching and learning. Finally, considering the empathic field, it evaluates the types of feedback that enhance the impact on learning outcomes. Understanding these topics is crucial for any context where new learning tools are being integrated. This SLR is a valuable resource for techno-pedagogical designers seeking to learn about empathic PCAs, as it covers each stage of the process. Furthermore, future research could draw from the ideas presented in this study while addressing the challenges discussed and exploring new research directions.

The remaining sections of this study are structured as follows. Second section details the methodology of the systematic review and its phases. Third section presents the findings of the study. Fourth section discusses the results, limitations and conclusions.

METHOD

The review process method is divided into several sections: protocol and registration, eligibility criteria, information sources, search, study selection and so on. Each section is further subdivided into several steps, each of which is described below.

Protocol and registration

This study explores existing literature on the design, implementation and evaluation of empathic PCA by adopting the Preferred Reporting Items for Systematic Reviews and Meta Analyses (PRISMA) framework (Moher et al., 2009) and following the Guidelines for performing Systematic Literature Reviews in Software Engineering (Kitchenham & Charters, 2007). A systematic review is "a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review" (Moher et al., 2009, p. 1). The systematic review registration number is osf.io/3xk6a.

Eligibility criteria

The primary aid of the search criteria was to investigate the latest advances in the design, implementation and evaluation of empathic conversational agents in education. To that effect, systematic reviews were preliminary searches to understand the study's context keywords and scope. The Population, Intervention, Comparison, Outcome, and Study (PICOS) method outlined by Petticrew and Roberts (2006) was used as a guideline to define the research directions. The study's population relates to the main keywords and their derivatives with similar connotations for the technical names of "conversational agents", "emotion" and "education" words. The intervention refers to the search context; keywords were used to filter studies that met the objective. All empathic PCA designs, implementations and evaluations were considered for comparison. The outcomes determined the data coding requirements and results, including empathic PCA design principles, set-ups for integrating students' previous knowledge, research approaches, assessment variables and feedback types that impact learning outcomes. Finally, the study designs were defined as experimental, action research or mixed methods.

Some essential criteria were defined to select articles for the review. Table 1 summarizes the inclusion and exclusion criteria applied for selecting the articles. First, the empirical studies included related to the design, implementation or evaluation of conversational agents, used in the educational field, and that consider the conversational agent's empathic

TABLE 1 Inclusion and exclusion criteria.

Inclusion criteria	Exclusion criteria
Must be an empirical study (quantitative, qualitative or mixed) on the conversational agents: it must mention the conversational agent's design, implementation and evaluation (all are mandatory)	Study on a conversational agent is nonempirical or uses the Wizard of Oz technique: it does not mention the conversational agent's design, implementation and evaluation
Must cover the practical application of a conversational agent in the teaching and learning process	The study does not consider the practical application of a conversational agent in the teaching and learning process
Must involve the conversational agent's empathic abilities in its design	The study does not consider the conversational agent's empathic abilities in its design
Must be a peer-reviewed conference paper or journal article	Book, book chapter, review, dissertation, article in press, abstract, poster or annotated bibliography
Must be the most recent paper on the subject	A previous paper presenting a conversational agent which is covered in a later article

TABLE 2 Quality assessment checklist.

No. of question	Question
Q1	Is the study relevant to the review?
Q2	Are the research aims and contributions identified?
Q3	Is the problem statement clear and does it integrate issues related to the review?
Q4	Is the experimental or action research set-up adequately described, and does it at least consider participants' learning performance and/or perceptions?
Q5	Are the methods/techniques clearly explained and analysed?
Q6	Are the results compared to previous studies/baseline on education and ICT?
Q7	Are sufficient data used for the evaluation of the empathic PCA?
Q8	Is the proposed empathic PCA evaluated using established criteria?
Q9	Is the conclusion explained clearly and linked to the purpose of the study?
Q10	Is the source of the article credible (published in a ranked venue)?
Q11	Has the article been cited in other publications?

abilities in its design. Second, only peer-reviewed conference papers and journal articles were included, excluding books, book chapters, reviews, dissertations, articles in the press, abstracts, posters or annotated bibliographies. Last, the most recent papers were included in the study; that is, previous papers mentioned in the newest articles were excluded.

To ensure a rigorous assessment of the articles included in the review, a quality assessment checklist consisting of 11 questions was developed and is presented in Table 2. Quality assessment is crucial in systematic reviews to ensure the validity of the results and reduce bias that may be caused due to the inclusion of less robust studies (Yang et al., 2021). The quality assessment also provides more detailed inclusion and exclusion criteria (Kitchenham & Charters, 2007). The elements essential to the data extraction and coding phases were considered, such as relevance to the research, clear identification of the research aims and validity of the results. Furthermore, the source's credibility was considered and evaluated using the ranking of the journal/conference and the number of citations of the study. These criteria were informed by principles of good practice for conducting empirical research in software engineering (Kitchenham et al., 2002).

Information sources

Various information sources were considered for the retrieval of relevant publications, ranging from general technology to enhanced learning topics. Accordingly, the research utilized the following nine digital databases: ACM Digital Library, IEEE Xplore, ProQuest, ScienceDirect, Scopus, SpringerLink, Taylor & Francis Online, Web of Science and Wiley Online Library.

Search

The articles were retrieved in November 2022. An extensive range of search strategies was used to retrieve the studies from the identified databases to raise the probability of identifying highly relevant studies. The logical operators 'AND' and 'OR' were used, combining the keywords identified in the planning process. Furthermore, the search was performed on the title, abstract and keywords to ensure that relevant studies were not left out. The following is the search query syntax used in all the identified databases: *("chatbot" OR "intelligent tutoring system" OR "smart personal assistant" OR "conversational agent" OR "conversational interface" OR "virtual agent" OR "digital agent") AND ("emotion" OR "emotional" OR "affective" OR "empathy" OR "empathic" OR "empathetic" OR "sentiment" OR "feeling") AND ("education" OR "learning" OR "teaching"*). Third, only articles published in English were included to eliminate the bias that may result from poor translation. The study period was determined to be from 1 January 2018 to 31 October 2022, as the development of agents with the integration of AI techniques has emerged in recent years, making 5 years sufficient to view research trends on empathic PCAs.

Study selection

In this phase, the inclusion and exclusion criteria were applied to screen the retrieved articles for eligibility following the PRISMA framework (Moher et al., 2009). This framework provides detailed guidelines and a structured approach to study selection. There were three steps in the study selection, and a pilot review was conducted to test the criteria and the protocol using a single database. First, records identified as duplicates through a computer supervised by a human were eliminated. Second, the inclusion and exclusion criteria were applied to ensure that only relevant papers would be included (title and abstract) and to assess relevance and eligibility. The study used only one screener in the eligibility phase. Finally, a quality assessment of the remaining articles was performed after the inclusion and exclusion criteria. The quality assessment was performed using the assessment checklist using 0 and 1, where 1 represents that the criteria are met and 0 represents not met. Only articles meeting at least 8 of the 11 criteria (ie, more than 70%) were included. One point was assigned to the paper having at least one citation to assess the number of citations. It should be noted that this quality assessment is a means of determining whether the selected paper is relevant to the contribution of this study but does not attempt to criticize any of the studies and their findings. As for data management, the sources obtained from the database searches and the authors' decisions were stored in an XLSX file. In addition, the studies included in the SLR were stored in Mendeley Reference Manager.

Data collection process

The method of data extraction from reports included in the SLR was in-depth reading. Some main details of the reports included were added to a matrix in the XLSX file, such as author(s), year, title, abstract, keywords, research questions, purpose, characteristics of the empathic PCA, methods, results, conclusions and research gaps. The next columns were according to the RQ. The words 'not mentioned' were recorded if the information was not found. The data items were design principles, set-ups for integrating students' previous knowledge, research approaches and variables for assessing the empathic PCA, and feedback types that impact learning outcomes. Two in-depth readings of the entire text were necessary to obtain and confirm data. Methods for assessing the risk of bias in individual studies were not used because the reports were evaluated by the quality assessment mentioned in Table 2.

Synthesis of results process

The methods of handling data and combining results of studies were a comparative analysis and synthesis of results recorded in the respective RQs' columns. No assessment of the risk of bias that may affect the cumulative evidence was used. The reason was the two in-depth readings of the entire text, which avoids the decontextualization of information-selective reporting within studies. Furthermore, to select the final sample, the screener discussed the quality assessment of the reports with the other authors (expert panel), who are experts on the link between emotion and Computer-Supported Collaborative Learning.

RESULTS

This section presents the results obtained from the SLR. 1162 results were retrieved, with the highest number of studies coming from Scopus because it is generic and sources publications from all domains. The results of the search are presented in Table 3.

After retrieving the search results, a bibliometric analysis of the results was performed to analyse the research areas. Figure 1 shows the visualization of the terms in the results, constructed using VOSviewer (Van Eck & Waltman, 2023). The diagram presents the significance of and interconnections between the frequently occurring terms extracted from the keywords search results. The size of the shape and the label associated with the term

TABLE 3 Search results.

Database	Search results
ACM Digital Library	37
IEEE Xplore	217
ProQuest	45
ScienceDirect	33
Scopus	429
SpringerLink	230
Taylor & Francis Online	1
Web of Science	165
Wiley Online Library	5

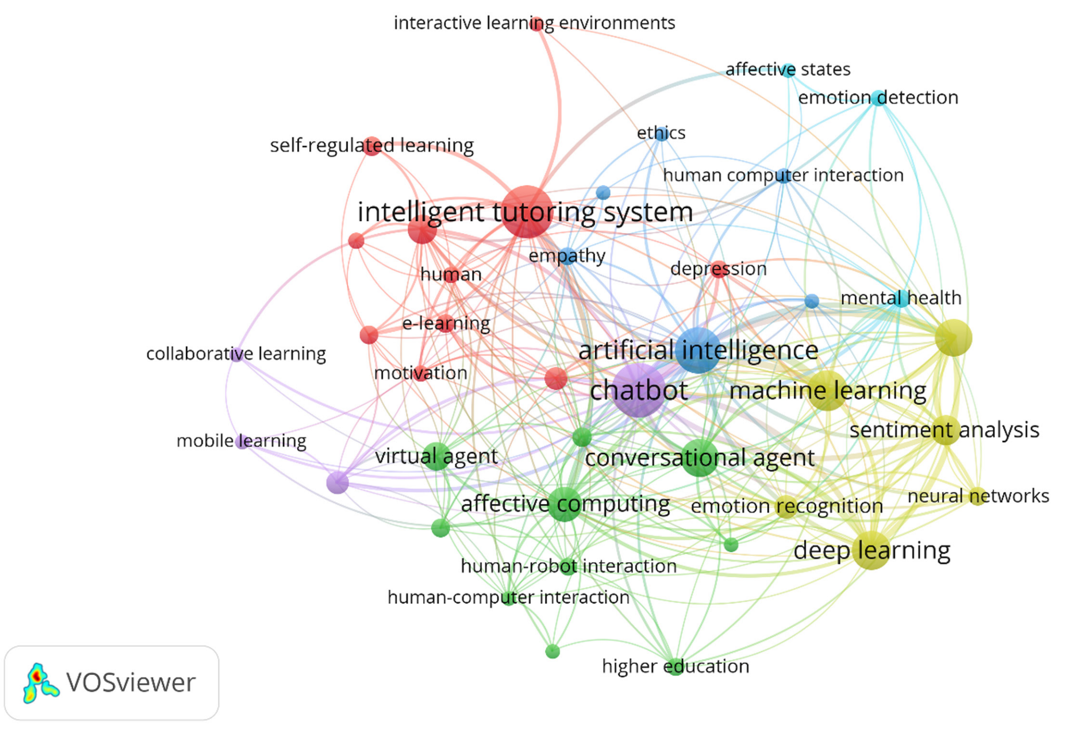

FIGURE 1 Bibliometric analysis of search results.

determines its importance. The colour of the terms determines the clusters in the visualization. Each cluster represents terms related to each other in that group, and the distance between the clusters represents the relatedness of the clusters.

The visualization of the terms in the extracted studies reveals several clusters. This shows that there are various dimensions of studies on empathic PCAs and current research trends from ITSs, affective computing, AI, machine learning (ML) and mental health. The clusters are tightly overlapped, indicating that several aspects of the studies are interrelated. Considering only the clusters with highly weighted terms, four main clusters can be seen in the visualization. The first and central cluster (purple) includes the following keywords: collaborative learning, education and mobile learning. This cluster shows that research is active in this area and related to empathic PCAs. The second cluster (blue) holds keywords such as empathy, ethics, human–computer interaction, intelligent virtual agent and virtual agent, meaning that the area of research in this cluster is AI. In the third cluster (red), the significant keywords are affect, depression, e-learning, embodied conversational agents (ECAs), emotion, human, interactive learning environment, learning and motivation, showing research relating more to ITSs. Finally, in the fourth, the significant keywords are Deep Learning, emotion recognition, NLP, Neural Networks and Sentiment Analysis, indicating that research in this cluster is about ML (aquamarine). Considering that there was a subsequent screening process, examining these clusters only provides a context of where the research finding is located for better analysis and discussion.

Figure 2 shows the number of studies identified, screened, assessed for eligibility and included in the review, with reasons for exclusion. Most of the articles ($n=740$) were excluded at this stage as they did not match the inclusion criteria. In some cases, a full-text review was necessary to check compliance with the inclusion and exclusion criteria. As discovered in the network analysis of the search terms, most of the articles evaluated the agent's

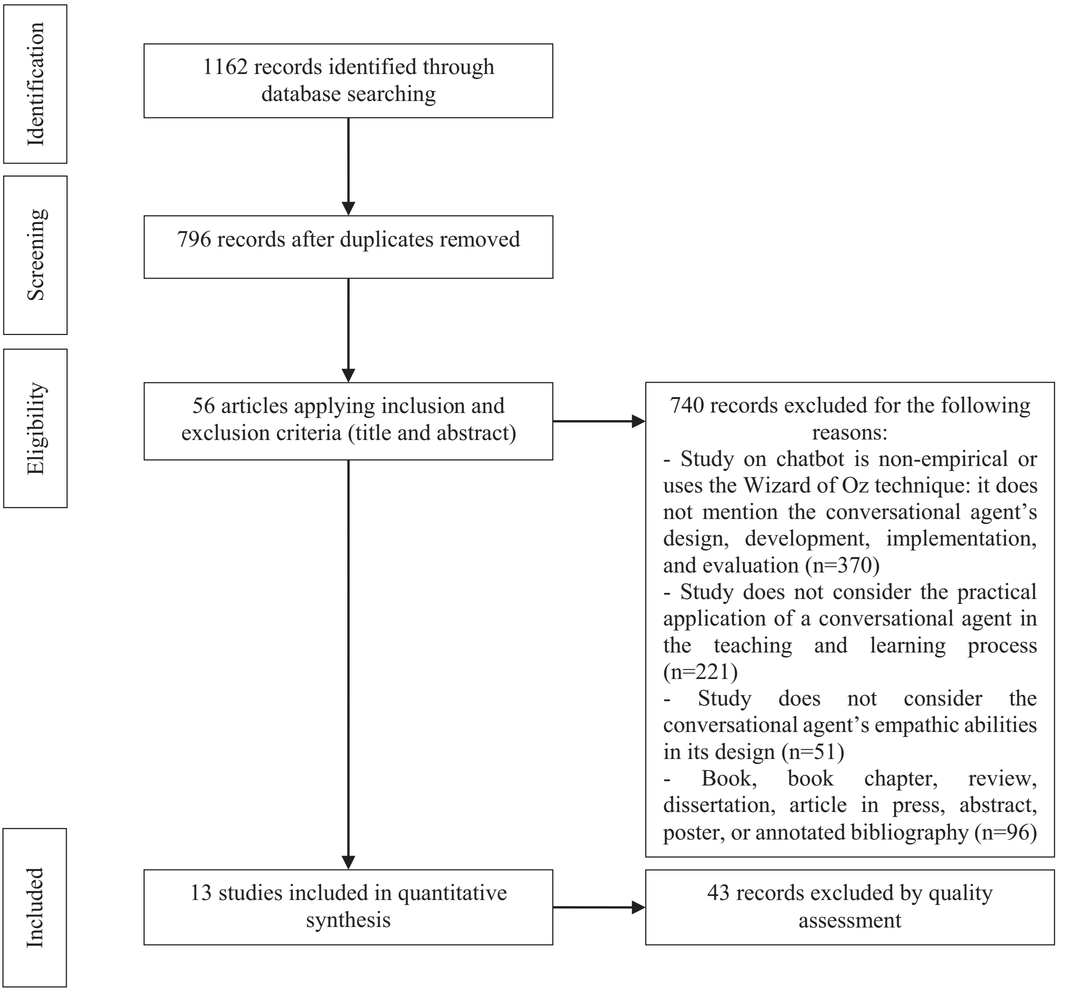

FIGURE 2 Flow of information in the systematic literature review phases.

empathic component without incorporating these abilities in its stage design. These articles were excluded as they did not contribute to the study. Finally, 13 studies were included in the SLR. Table 4 presents a detailed quality assessment of the articles included, showing that all included articles are of excellent quality. The expert panel also conducted this assessment and agreed with the evaluation report and its results.

Study characteristics

Data analysis was conducted on all the relevant features identified in the planning phase. Metadata analysis includes a variety of data to answer the RQs. Five themes make up the main data corpus on empathic PCAs. These are design principles, set-ups for integrating the students' previous knowledge, research approaches, variables for assessing the empathic PCA and feedback types that impact learning outcomes. Supplementary material (Table S1) shows the main characteristics of the studies.

TABLE 4 Quality assessment.

Study ID	Author(s), year	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Total	%
S1	Arguedas and Daradoumis (2021)	1	1	1	1	1	1	1	1	1	1	1	11	100
S2	Ayedoun et al. (2020)	1	1	1	1	1	1	1	1	1	1	1	11	100
S3	Jimenez et al. (2018)	1	1	1	1	1	1	1	1	1	1	0	10	91
S4	Kumar (2021)	1	1	1	1	1	1	1	1	1	1	1	11	100
S5	Liu et al. (2022)	1	1	1	1	1	1	1	1	1	1	0	10	91
S6	Long et al. (2019)	1	1	1	1	1	1	1	1	1	1	0	10	91
S7	Munshi et al. (2018)	1	1	1	1	1	1	1	1	1	1	1	11	100
S8	Oker et al. (2020)	0	1	1	1	1	1	1	1	1	1	1	10	91
S9	Santos et al. (2020)	1	1	1	1	1	1	1	1	1	1	1	11	100
S10	Scholten et al. (2019)	0	1	1	1	1	0	1	1	1	1	1	9	82
S11	Terzidou et al. (2018)	1	1	0	1	1	1	1	1	1	1	1	10	91
S12	Wambsganss et al. (2021)	1	1	1	1	1	1	1	1	1	1	1	11	100
S13	Wu et al. (2020)	1	1	1	1	1	1	1	1	1	1	1	11	100

Synthesis of results

Corresponding to the report's number, 13 empathic PCAs were found. Supplementary material (Table S2) shows an overview of each agent. Each one has a name, domain and empathic abilities. Moreover, each one corresponds to a type and was applied to an education level.

RQ1. Design principles

In the selected articles, it was possible to identify five sets of empathic PCA construction criteria (see Table 5). The authors presented their views on the design principles, including reliability, interpersonal communication, learning, experience, web-based application, rich media content, individual empathy feedback, theory-based learning, comparison with peers, inputting natural text and differentiated feedback, ontology elements, communication strategies and dialogue moves. These design principles aim to provide education, support and feedback in an empathic manner. Despite some design principles being focused on specific learning tools, all share the same functionalities as empathic PCAs.

RQ2. Set-ups for integrating the students' previous knowledge

This SLR reveals a lack of clarity on empathic PCA set-ups that integrate the previous knowledge of the students. Arguedas and Daradoumis (2021) provide some potential solutions by

TABLE 5 Empathic PCA design principles.

Study ID	Empathic PCA design principles
S2	The authors discussed Communication Strategies such as simplification, asking for clarification and suggesting answer patterns. They also discussed the Affective Backchannel: encouraging, sympathetic and reassuring
S3	Three ontology elements are used in Intelligent Tutor for Object-Oriented Programming (TIPOO, by its acronym in Spanish), an ITS agent with the same capabilities as empathic PCA. TIPOO is based on a dialogue module, provides individual support with a friendly attitude and includes affectivity to improve student motivation
S4	Four design principles for empathic PCA: reliability, interpersonal communication, learning and experience. The chatbot was designed to be accessible through Mobily instant messaging using Telegram, with privacy features for the data shared. It also mimics interpersonal communication and provides specific learning content for the subject. The agent was designed to promote active learning, reflection, metacognition and communication through notifications, content and guidelines. The agent also has a human-like interaction, with affective interaction, greetings and empathy
S9	The authors discussed the conversation flow based on Gottman's Emotion Coaching and four dialogue moves: feedback, pumping, confirming and reflecting
S12	There are six design principles for the adaptive learning tools' designers. The chatbot's empathic abilities are cross-cutting in the design principles: the design should be based on a web-based application with a responsive user interface and an intuitive user experience, rich media content, an individual empathy feedback mechanism, a theory-based learning scenario, a comparison of empathy skills with peers, and a function for inputting natural text and giving differentiated feedback

Note: S2 and S9 present two sets of communication strategies and dialogue moves respectively, that have a similar function in the design of empathic PCA.

focusing on students' previous work in virtual workspaces, such as wikis, blogs and forums. They use the Moodle environment to transmit dimensional and categorical emotional information to the Affective Pedagogical Tutor (APT), which provides cognitive and affective feedback. However, the authors do not consider the dialogue between the agent and the students as a source of information for future sessions. On the contrary, Wu et al. (2020) implemented a chatbot that saves the dialogue between the user and the bot (user's message and chatbot response) in a database only for future improvement. Another approach of other authors is to administer a pretest to assess the student's competencies, then configure the agent for the entire group based on the results. However, the approach by Arguedas and Daradoumis (2021), Wu et al. (2020) and other authors does not consider the previous conversation with the agent to update its database on the learner. The research approaches are discussed in the following section.

RQ3. Research approaches

Empathic PCA has been the subject of a range of research studies, each using different research designs. Table 6 shows the research approaches used in the agents' implementation. Nine of the studies employ a quantitative research approach, while four utilize mixed research combining quantitative and qualitative approaches. Seven studies used a quasi-experimental research design because these are set up in predefined groups, either control or experimental group(s). The other uses an experimental design, except for the case of Wu et al. (2020), who does not express how the participants' recruitment process was carried out. All the studies have a posttest component, while only just over half (7 out of 13) have a pretest component. Control groups are present in eight of the studies, while five lack control groups. The number of experimental groups ranges from one to six, with most studies having either a single experimental group (9 out of 13) or two experimental groups (2 out of 13). These findings highlight the diversity of approaches used in the research of empathic PCA, with a focus on both quantitative and mixed methods, and the use of pre and posttesting.

RQ4. Variables for assessing the empathic PCA

The authors place special emphasis on the learning outcomes when they evaluate the learning articulated by the empathic PCA. Based on the literature, learning outcomes are not only what students should know and be able to do at the end of a course or programme, but also how they perceive the learning process itself. These outcomes consist of two variables: learning performance and student perceptions. To determine these outcomes, data collection instruments are applied at the end of the implementation process and, in most cases, are compared to the students' initial state and/or the control group.

Learning performance

Several authors, including Kumar (2021), Long et al. (2019), Munshi et al. (2018) and Oker et al. (2020), use tests to evaluate learning performance in terms of content, procedures or attitudes. For example, Kumar (2021) assesses student project development using a test, and Long et al. (2019) use a test to evaluate the learning performance of research method knowledge articulated by their multiagents. Meanwhile, Munshi et al. (2018) and Oker et al. (2020) utilize tests to evaluate science domain concepts and understanding of causal relations and numeracy exercises respectively.

TABLE 6 Research approaches used in the implementation.

Research approach	ID	Control group	Experimental group	Research design	Pretest	Posttest	Learning performance	Student perceptions
Quantitative	S1	1	1	Q-E		x		x
	S2	0	6	E	x	x		x
	S3	0	1	Q-E		x		x
	S6	1	2	E	x	x	x	
	S7	0	1	Q-E	x	x	x	
	S8	1	1	E		x	x	x
	S10	1	3	E		x		x
	S11	0	1	Q-E		x		x
	S13	2	1	U		x		x
Mixed	S4	1	1	Q-E	x	x	x	x
	S5	1	2	Q-E	x	x		x
	S9	0	1	Q-E	x	x		x
	S12	1	1	E	x	x		x

Note: Q-E = Quasi-experimental; E = Experimental; U = Unidentified. S3 writes 'qualitative experiment' as research design; however, its procedure and results are quantitative.

Student perceptions

The authors have used various data collection techniques to assess student perceptions (see Table 7).

Other data sources: Conversational logs

The effectiveness of dialogic learning in determining the learning outcomes articulated by a chatbot was evaluated by Santos et al. (2020) using an observation checklist and by Liu et al. (2022) using deductive coding according to a story structure: characters, actions, consequences and causal relationship. In the first case, the instrument applied to the conversation logs allowed them to assess the chatbot's ability to accurately detect emotions in the child's input and challenges in communication. In the second case, the deductive coding applied to the conversation logs allowed them to determine how students interacted with the chatbot and perceived social connection. In this regard, dialogic learning plays a crucial role in validating acquired learning and demonstrating retention of emotions in the conversation log.

The reports have considered two variables when assessing empathic PCA learning outcomes: learning performance and student perceptions (see Figure 3). Learning performance is linked to the domain and objective of the agent, including content, procedures or attitudes. Student perceptions are complex and involve many dimensions. Figure 3 shows the dimensions to evaluate the student perceptions that were found in the SLR and are frequently mentioned, leaving the unpopular in ellipses. One of the most relevant is the affective bond. Other similar dimensions between the studies are interaction enjoyment and confidence perception. This SLR highlights the importance of differentiating between individual and teamwork perceptions. Considering the previous section, both variables are evaluated using the quantitative approach, but only the student perceptions integrate the qualitative approach in mixed-method research.

RQ5. Feedback types that impact learning outcomes

To analyse the types of empathic PCA feedback that impact learning outcomes, the next section divides this into two variables: learning performance and student perceptions. The variables are influenced positively according to the feedback types and appear to be positively correlated (see Figure 3).

Learning performance

According to Kumar (2021), the design principles of a chatbot play a crucial role in determining its effectiveness in positively impacting learning performance. The design principles are mentioned in Table 5. Long et al. (2019) discussed the impact of Multiagent Intelligent Tutoring System feedback on student learning performance. The authors found that the agents' cognitive support positively impacts students with low rejection sensitivity in confusion regulation. On the other hand, the empathic support of the agents positively impacts students with high rejection sensitivity in confusion regulation. Munshi et al. (2018) discussed the impact of conversational agent feedback on student learning performance. The authors found that high-scoring students were delighted when the mentor agent provided hints, analysed their progress and praised their progress towards generating the correct map. Oker et al. (2020) discuss the impact of virtual tutor feedback on student learning performance. The authors found that students who experimented with the bimodal condition had both longer reaction times representing engagement and more correct responses, whereas this was not the case in the unimodal condition.

Empathic PCAs can provide distinct types of feedback that positively impact student learning performance. These feedback types are configured in the design principles and

TABLE 7 Data collection techniques to assess student perceptions.

Instrument(s)	Instrument quality	Objective	Study ID
A questionnaire	R	To assess cognitive and affective feedback and students' emotional states	S1
A questionnaire	V, R	To evaluate student perceptions on three dimensions: confidence, anxiety and desire to communicate	S2
A survey	–	To gauge students' system preferences	
A questionnaire	V, R	To report students' motivation dimension based on their perceptions of the TIPOO	S3
Questionnaires		To analyse the affective-motivational perceptions of the students, including cognition, creative self-efficacy, motivational belief, and learning perception, and a questionnaire on teamwork perceptions	S4
1. Need for Cognition Scale–6	V, R		
2. Creative Self-Efficacy	V		
3. Motivated Strategies for Learning Questionnaire (MSLQ) and modified MSLQ	V, R		
4. Perception of learning	–		
5. Team Assessment Survey Questions	–		
Questionnaires		To understand student perceptions of the learning environment	S5
1. A flow perceptions questionnaire	V, R		
2. Godspeed questionnaire	V, R		
3. Situational interest questionnaire	V, R		
An interview	–	To understand the implications of the quantitative data and to further explore students' reactions to the chatbot	
Social Support Questionnaire for Children	V, R	To evaluate bimodal and unimodal agent empathy	S8
An interview	V	To determine the child's experience in storytelling and diary writing and their familiarity with a chatbot	S9
A survey	V	To determine the chatbot's ability to engage in a conversation that feels natural to the child (humanity metric) and the child's enjoyment and perceptions on the chatbot (affect metric)	

TABLE 7 (Continued)

Instrument(s)	Instrument quality	Objective	Study ID
Questionnaires		To assess feedback and autonomy, attention and relevance, involvement and rapport perceptions	S10
1. EGameFlow	V, R		
2. Instructional Materials Motivation Survey	V, R		
3. Personal Involvement Inventory	V, R		
4. Rapport scale	V, R		
A questionnaire	–	To measure general perceptions of the case study	S11
A survey (open and closed questions)	V, R	To evaluate perceived usefulness, intention to use, ease of use, level of enjoyment and design principle perceptions	S12
A questionnaire	–	To assess feelings of isolation and detachment, course-related quality assurance performance and user experience to compare the chatbot with teacher counselling services provided by the e-learning platform	S13

Note: V = validated; R = reliable.

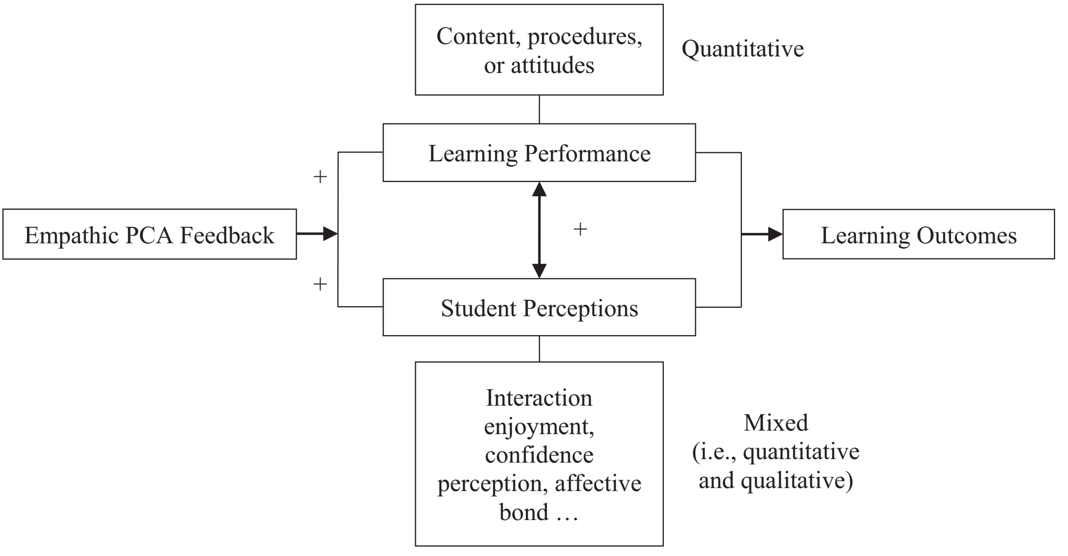

FIGURE 3 Framework to evaluate the learning outcomes of empathic Pedagogical Conversational Agents. "+" = positive influence; "→" = direction of influence; "↔" = bidirectional influence; "–" = include.

include cognitive and empathic, hints and bimodal feedback. Likewise, feedback on analysis and praise of student progress has a positive impact. In this regard, it is important to understand the different types of empathic PCA feedback can provide and how to design them effectively.

Student perceptions

This section compares the different types of empathic agent feedback that positively influence student perceptions.

Arguedas and Daradoumis (2021) find that cognitive and affective feedback from the APT has a positive impact on student perceptions. Cognitive feedback involves encouraging student proposals, informing students about the learning activity and arousing student interest in the topics. On the other hand, affective feedback involves fostering a creative environment, giving students confidence and motivating students to think that the lesson goals are achievable and to become more involved in the learning activities. Ayedoun et al. (2020) find that Communication Strategies (CS) and Affective Backchannel (AB) have a positive impact on student perceptions. The authors observed that learners with a lower willingness to communicate tend to prefer AB over CS, while their counterparts with higher willingness to communicate tend to favour CS over AB. CS includes simplification, asking for clarification and suggesting an answer pattern, while AB includes encouraging, sympathetic and reassuring feedback.

Jimenez et al. (2018) find that affective dialogue, based on encouragement phrases, has a positive impact on the motivation of students with low academic performance, female students and engineering students. The authors concluded that affective feedback significantly impacts motivation, particularly in these cases. Kumar (2021) finds that chatbot feedback does not significantly affect individual student perceptions. However, the author finds that the chatbot was positively received in teamwork perceptions due to its design for collaborative learning. The chatbot design principles are mentioned in Table 5. Liu et al. (2022) find that chatbots with book-talk abilities and social affective cues have a positive impact on student perceptions. Chatbot's book talk is an interactive conversation or discussion about a book that takes place between a human reader and a chatbot. The chatbot's book-talk capabilities included an invitation, a storyline question, acknowledge, repeat, follow-up and expansion suggestions. The chatbot's empathic abilities included praise, asking for feelings, feeling sharing and recommendations.

Oker et al. (2020) find that students had positive perceptions when interacting with an empathic agent that displayed coherent facial expressions. This result suggests that the use of nonverbal cues, such as facial expressions, can increase students' motivation and engagement. On the other hand, Santos et al. (2020) suggest that positive student perceptions of chatbots are related to the design principles of conversation flow and dialogue moves. The design principles are mentioned in Table 5. Scholten et al. (2019) find that ECA characteristics positively affect student perceptions of feedback and autonomy, including visibility and speech communication. The authors find that a visible agent positively affected feedback and autonomy, and that speech communication by the agent positively affected feedback, regardless of gender. The authors also find a gender effect, where male participants rated the visible agent higher than female participants and rated the nonvisible agent lower than female participants.

Terzidou et al. (2018) find that empathic PCA feedback that helps students control their heart rate in cases of high anxiety has a positive impact on their perceptions. This finding highlights the importance of agents providing support and guidance to students in stressful situations. Wambsganss et al. (2021) find that the design principles of chatbot feedback have a positive impact on student perceptions. The design principles are listed in Table 5. Wu et al. (2020) find that chatbot feedback that is relaxing, informative, entertaining and suited to the popular culture of K–12 students positively impacts student perceptions. This finding emphasizes the importance of considering the context and tone of feedback when designing agents.

Feedback types have a significant impact on student perceptions, and the design of empathic PCA feedback must be well thought out and tailored to meet the specific needs of the learners. Whether through cognitive and affective feedback, scaffold design, chatbot's book talk and social affective cues, coherent facial expressions, ECA characteristics, support for students with significant levels of anxiety, or popular culture topics, empathic PCAs have the potential to enhance student learning outcomes by shaping positive student perceptions. These findings highlight the importance of considering numerous factors when designing feedback types to positively impact student perceptions and enhance the learning experience.

DISCUSSION

This section presents a summary of evidence from the 13 empathic PCAs, limitations and conclusions. Some of the statements in this summary were also presented as conference proceeding (Ortega-Ochoa, 2023). It is structured in five points based on RQs.

First, five sets of design principles for empathic PCAs were found in the review (Ayedoun et al., 2020; Jimenez et al., 2018; Kumar, 2021; Santos et al., 2020; Wambsganss et al., 2021). Each has its particularities. Nevertheless, four elements are highlighted in the building of an empathic PCA. These are the transversality of empathic abilities (ie, the presence of this design principle in all agents), the promotion of dialogic learning (communication and learning), proficiency in the field of knowledge and personalized feedback according to the student's level. Furthermore, two reports mention that the reliability of the environment and user experience are also principles to consider in the design (Kumar, 2021; Wambsganss et al., 2021).

Second, some authors propose set-ups to integrate the student's previous knowledge in the interaction with the empathic PCA; for instance, dialogues in the Learning Management System (Arguedas & Daradoumis, 2021) or tests for the students. These are rich data sources for ascertaining the students' state, but this is not the only one. Various authors mention the analysis of conversational logs at the end of the intervention (Wu et al., 2020). However, in the reports, there is a lack of automatic integration of the conversational logs as a data source for future interaction with the agent in the same session and experiment.

Third, the research design preferred by researchers is experimenting with empathic PCA implementation. In this case, the quantitative approach is predominant (see Table 6), although four studies utilized mixed method designs for a comprehensive understanding of the effectiveness of the learning tool (Kumar, 2021; Liu et al., 2022; Santos et al., 2020; Wambsganss et al., 2021). These have chatbots as conversational agent types. This method of implementation makes its evaluation feasible, specifically, in the last part of the intervention, owing to all reports having a posttest. Only seven have a pretest, making comparison with the student's initial state impossible in the other studies. However, four of them had a control group rather than a pretest as a resource.

Fourth, the authors consider two variables to evaluate the effectiveness of the learning articulated by the empathic PCA. Learning performance (1) refers to content, procedures or attitudes (Kumar, 2021; Long et al., 2019; Munshi et al., 2018; Oker et al., 2020). Tests are the preferred instrument for data collection. The content of the test will depend on the domain and objective relating to the empathic PCA. The quantitative approach is the only evaluation method applied. As for student perceptions, (2) many dimensions are considered. The affective bond dimension is evaluated by most reports. Other dimensions are interaction enjoyment (eg, Santos et al., 2020; Wambsganss et al., 2021; among others), confidence perception (eg, Ayedoun et al., 2020; Liu et al., 2022; among others) and so forth. If the chatbot was designed with empathic and teamwork skills, this dimension must also be evaluated (Kumar, 2021). The main instruments are questionnaires, surveys and interviews. The indicators are quantitative, although there are also open questions. Our results are broadly consistent with previous research such as the one presented by Fitrianie et al. (2019) on Intelligent Virtual Agents, where there is no reuse of instruments even in the evaluation of empathic PCA, while the trend towards the study of new dimensions is maintained.

Fifth, the feedback types defined in the design principle play a relevant role in achieving positive results (Kumar, 2021) in both learning performance and student perceptions (Santos et al., 2020; Wambsganss et al., 2021). In general, these variables seem to be positively correlated (Kumar, 2021; Oker et al., 2020). On the one hand, cognitive and empathic (Long et al., 2019), hints (Munshi et al., 2018), and bimodal feedback (Oker et al., 2020) are necessary for positive learning performance. Likewise, the analysis and praise of student progress have a positive impact (Munshi et al., 2018). On the other hand, cognitive (Arguedas & Daradoumis, 2021) and affective feedback (Arguedas & Daradoumis, 2021; Jimenez et al., 2018), scaffold design (Ayedoun et al., 2020), chatbot's book talk and social affective cues (Liu et al., 2022), coherent facial expressions (Oker et al., 2020), ECA characteristics (Scholten et al., 2019), support for students with significant levels of anxiety (Terzidou et al., 2018) or popular culture topics (Wu et al., 2020) are necessary for positive student perceptions. Nevertheless, the particularities of each student, for instance, gender (Scholten et al., 2019), will influence the results. Overall, most of the feedback types share similar characteristics, only some feedback types affect one variable more than the other. Figure 3 presents the resulting SLR framework for evaluating the learning outcomes of empathic PCA.

Limitations

This study has three main limitations and the research gaps are accordingly. First, the time range for literature selection was limited, which may not have been sufficient to generate a framework that adequately assesses the learning outcomes articulated by the empathic PCA. While the study does consider the latest advances in the field, a more comprehensive SLR is needed to validate the theoretical results. Additionally, a practical evaluation is necessary to better understand the potential of these agents. Second, it is necessary to research the level of integration of the empathic field in the agents, for example, the exclusive identification of emotion or feedback considering the current emotional state of the learner. Regarding the latter, a framework is needed that provides feedback in view of the affective and cognitive state. Third, the lack of a detailed report on the development stage of the empathic PCA presented in this review was due to the vast diversity of constructions and the amplitude of the results. One of the key characteristics of this SLR is its consideration of all types of PCA that can integrate empathic abilities (chatbots and ITS agents). Future research could benefit from a more comprehensive and detailed investigation of the comparison of each agent's type mentioned above, specifically in its development stage.

Conclusions

This SLR described the empathic PCA's design, implementation and evaluation stages through the identification of 13 learning tools, which represent a relevant contribution to the technology-enhanced learning field.

We may draw some conclusions on the design and implementation stage of empathic PCAs. First, the design principles of most of these agents are the transversality of empathic abilities, the promotion of dialogic learning, proficiency in the field of knowledge and personalized feedback according to the student's level. Based on the results, the design principle of the transversality of empathic abilities should not be isolated, but cross-cutting among the other principles. In general, these design principles form the basis for defining feedback types. Second, there is a lack of clarity in integrating previous agent and learner conversations in a database to determine learning states and personalize responses during the same session. Finally, most agents' implementations have a quantitative approach. Here, detailed data must be added for a comprehensive evaluation.

As for the evaluation stage of the empathic PCA, two variables are evaluated to determine the learning outcomes articulated by the agents: learning performance and student perceptions. The results suggest that different feedback types have a positive impact on both variables, indicating a correlation between them. Therefore, the study proposes a preliminary framework for evaluating the learning outcomes of empathic agents, which requires further theoretical research and practical validation to establish its validity. For instance, the framework and each theme can be compared with other SLRs that accomplish the same objective.

While some future directions for researchers have been outlined in the limitation section, it is relevant to address certain ethical considerations for both researchers and practitioners. In particular, the substantial collection of data, primarily from students, and its subsequent handling by the algorithm raise concerns about the ethical implications of agents operating beyond scheduled learning times. Questions arise regarding whether such practices align with the preferences and expectations of contemporary learners as a necessary pedagogical strategy. Relevant research in this area is the work of Baker and Hawn (2022) and Kizilcec and Lee (2022). The former provides an overview of the sources of algorithmic bias along the machine learning pipeline in greater depth and the latter discusses techniques ranging from measurement to model learning to action in an algorithm system that can be adopted to improve algorithmic fairness in education. Additionally, there is a unanimous call in the literature for a design method to develop a more consolidated empathic PCA.

Various proposals advocate for the adoption of Responsible AI principles, such as the Value Sensitive Design—a well-established approach that systematically incorporates human values throughout the technology design process (Friedman & Hendry, 2019). These principles can be applied to enhance the empathic capabilities of PCAs. The timeline for realizing these advancements remains uncertain, but it is undeniable that AI will profoundly reshape multiple facets of life and society in the forthcoming decades.

ACKNOWLEDGEMENTS
Not applicable.

FUNDING INFORMATION
This work was supported by the Universitat Oberta de Catalunya.

CONFLICT OF INTEREST STATEMENT
We have no conflicts of interest to declare that are relevant to the content of this article.

DATA AVAILABILITY STATEMENT
All data generated or analysed during this study are included in this published article. The search keywords and databases used in the SLR are provided in the paper. Supplementary material (Table S1) lists all the papers included in the SLR. Furthermore, the data encoding of the papers analysed during the current study is available from the corresponding author upon reasonable request.

ETHICS STATEMENT
Since this study had not involved human subjects as participants, no approval was needed from an institutional ethics committee.

ORCID
Elvis Ortega-Ochoa https://orcid.org/0000-0003-4634-6288
Marta Arguedas https://orcid.org/0000-0002-6609-122X
Thanasis Daradoumis https://orcid.org/0000-0001-9362-4469

REFERENCES
*Arguedas, M., & Daradoumis, T. (2021). Analysing the role of a pedagogical agent in psychological and cognitive preparatory activities. *Journal of Computer Assisted Learning*, *37*, 1167–1180. https://doi.org/10.1111/jcal.12556

*Ayedoun, E., Hayashi, Y., & Seta, K. (2020). Toward personalized scaffolding and fading of motivational support in L2 learner-dialogue agent interactions: An exploratory study. *IEEE Transactions on Learning Technologies*, *13*(3), 604–616. https://doi.org/10.1109/TLT.2020.2989776

Baker, R. S., & Hawn, A. (2022). Algorithmic bias in education. *International Journal of Artificial Intelligence in Education*, *32*, 1052–1092. https://doi.org/10.1007/s40593-021-00285-9

Bilquise, G., Ibrahim, S., & Shaalan, K. (2022). Emotionally intelligent Chatbots: A systematic literature review. *Human Behavior and Emerging Technologies*, *2022*, 1–23. https://doi.org/10.1155/2022/9601630

Fitrianie, S., Bruijnes, M., Richards, D., Abdulrahman, A., & Brinkman, W.-P. (2019). What are we measuring anyway?—A literature survey of questionnaires used in studies reported in the intelligent virtual agent conferences. In C. Pelachaud & J.-C. Martin (Chairs), *IVA '19: Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents* (pp. 159–161). Association for Computing Machinery. https://doi.org/10.1145/3308532.3329421

Friedman, B., & Hendry, D. G. (2019). *Value sensitive design: Shaping technology with moral imagination*. The MIT Press.

Hall, L., & Woods, S. (2006). The importance of similarity in empathic interaction. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 303–310). Idea Group Reference. https://doi.org/10.4018/978-1-59140-562-7

Huang, W., Hew, K. F., & Fryer, L. K. (2021). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. *Journal of Computer Assisted Learning*, *38*(1), 237–257. https://doi.org/10.1111/jcal.12610

*Jimenez, S., Juárez-Ramírez, R., Castillo, V. H., & Ramírez-Noriega, A. (2018). Integrating affective learning into intelligent tutoring systems. *Universal Access in the Information Society*, *17*, 679–692. https://doi.org/10.1007/s10209-017-0524-1

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, *11*, 47–78.

Kitchenham, B. A., & Charters, S. (2007). *Guidelines for performing systematic literature reviews in software engineering* [EBSE Technical Report, EBSE-2007-01]. Keele University and University of Durham.

Kitchenham, B. A., Pfleeger, S. L., Pickard, L. M., Jones, P. W., Hoaglin, D. C., El Emam, K., & Rosenberg, J. (2002). Preliminary guidelines for empirical research in software engineering. *IEEE Transactions on Software Engineering*, *28*(8), 721–734. https://doi.org/10.1109/TSE.2002.1027796

Kizilcec, R. F., & Lee, H. (2022). Algorithmic fairness in education. In W. Homes & K. Porayska-Pomsta (Eds.), *The ethics of artificial intelligence in education: Practices, challenges, and debates* (pp. 174–202). Routledge. https://doi.org/10.4324/9780429329067

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2022). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*, 973–1018. https://doi.org/10.1007/s10639-022-11177-3

*Kumar, J. A. (2021). Educational chatbots for project-based learning: Investigating learning outcomes for a team-based design course. *International Journal of Educational Technology in Higher Education*, *18*(65), 1–28. https://doi.org/10.1186/s41239-021-00302-w

Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., & Abraham, A. (2022). AI-based conversational agents: A scoping review from technologies to future directions. *IEEE Access*, *10*, 92337–92356. https://doi.org/10.1109/ACCESS.2022.3201144

*Liu, C.-C., Liao, M.-G., Chang, C.-H., & Lin, H.-M. (2022). An analysis of children' interaction with an AI chatbot and its impact on their interest in reading. *Computers & Education*, *189*, 1–16. https://doi.org/10.1016/j.compedu.2022.104576

*Long, Z., Luo, D., Gao, H., & Hu, X. (2019). The effects of rejection sensitivity on confusion regulation during learning in multiagent intelligent tutoring system environments. In A. M. Sinatra & J. A. DeFalco (Chairs), *Approaches and challenges in team tutoring workshop* (pp. 77–86). 20th Artificial Intelligence in Education Conference (AIED 2019). https://ceur-ws.org/Vol-2501/paper9.pdf

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, *6*(7), 1–6. https://doi.org/10.1371/journal.pmed.1000097

*Munshi, A., Rajendran, R., Ocumpaugh, J., Biswas, G., Baker, R. S., & Paquette, L. (2018). Modeling Learners' cognitive and affective states to scaffold SRL in open-ended learning environments. In T. Mitrovic & J. Zhang (Chairs), *UMAP '18: Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization* (pp. 131–138). Association for Computing Machinery. https://doi.org/10.1145/3209219.3209241

*Oker, A., Pecune, F., & Declercq, C. (2020). Virtual tutor and pupil interaction: A study of empathic feedback as extrinsic motivation for learning. *Education and Information Technologies*, *25*, 3643–3658. https://doi.org/10.1007/s10639-020-10123-5

Okonkwo, C. W., & Ade-Ibijola, A. (2021). Chatbots applications in education: A systematic review. *Computers and Education: Artificial Intelligence*, *2*, 100033. https://doi.org/10.1016/j.caeai.2021.100033

OpenAI. (2022, November 30). *ChatGPT: Optimizing language models for dialogue*. OpenAI, L.L.C. https://openai.com/blog/chatgpt

Ortega-Ochoa, E. (2023). An empathic pedagogical conversational agent for development of competencies: A research plan. In D. Di Mitri, A. Ortega-Arranz, & O. Poquet (Eds.), *Proceedings of the Doctoral Consortium of the 18th European Conference on Technology Enhanced Learning co-located with the 18th European Conference on Technology Enhanced Learning (EC-TEL 2023)* (pp. 18–25). Doctoral Consortium of the 18th European Conference on Technology Enhanced Learning. https://ceur-ws.org/Vol-3539/paper3.pdf

Petticrew, M., & Roberts, H. (2006). *Systematic reviews in the social sciences: A practical guide*. Wiley-Blackwell.

Prendinger, H., & Ishizuka, M. (2005). The empathic companion: A character-based Interface that addresses users' affective states. *Applied Artificial Intelligence*, *19*, 267–285. https://doi.org/10.1080/08839510590910174

Rapp, A., Curti, L., & Boldi, A. (2021). The human side of human chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human Computer Studies*, *151*, 102630. https://doi.org/10.1016/j.ijhcs.2021.102630

Research and Markets. (2022, July 1). *Size of the chatbot market worldwide from 2021 to 2030 (in million U.S. dollars)* [Graph]. Statista. Retrieved February 20, 2023, from https://www.statista.com/statistics/656596/worldwide-chatbot-market

*Santos, K.-A., Ong, E., & Resurreccion, R. (2020). Therapist vibe: Children's expressions of their emotions through storytelling with a Chatbot. In E. Rubegni & A. Vasalou (Chairs), *IDC '20: Proceedings of the Interaction Design and Children Conference* (pp. 483–494). Association for Computing Machinery. https://doi.org/10.1145/3392063.3394405

*Scholten, M. R., Kelders, S. M., & Van Gemert-Pijnen, J. E. W. C. (2019). An empirical study of a pedagogical agent as an adjunct to an eHealth self-management intervention: What modalities does it need to successfully support and motivate users? *Frontiers in Psychology*, *10*, 1–14. https://doi.org/10.3389/fpsyg.2019.01063

Tegos, S., & Demetriadis, S. (2017). Conversational agents improve peer learning through building on prior knowledge. *Journal of Educational Technology & Society*, *20*(1), 99–111. https://www.jstor.org/stable/jeductechsoci.20.1.99

*Terzidou, T., Tsiatsos, T., & Apostolidis, H. (2018). Architecture and interaction protocol for pedagogical-empathic agents in 3D virtual learning environments. *Multimedia Tools and Applications*, *77*, 27661–27684. https://doi.org/10.1007/s11042-018-5942-4

Van Eck, N. J., & Waltman, L. (2023). *VOSviewer Manual* [version 1.6.19]. https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.19.pdf

*Wambsganss, T., Weber F., & Söllner, M. (2021). Designing an adaptive empathy learning tool. In T. X. Bui (Chair), *Proceedings of the 54th Hawaii International Conference on System Sciences* (pp. 54–63). University of Hawaii at Manoa. https://doi.org/10.24251/HICSS.2021.007

Weizenbaum, J. (1966). ELIZA—A Computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

*Wu, E.H.-K., Lin, C.-H., Ou, Y.-Y., Liu, C.-Z., Wang, W.-K., & Chao, C.-Y. (2020). Advantages and constraints of a hybrid model K-12 E-learning assistant Chatbot. *IEEE Access*, *8*, 77788–77801. https://doi.org/10.1109/ACCESS.2020.2988252

Yang, L., Zhang, H., Shen, H., Huang, X., Zhou, X., Rong, G., & Shao, D. (2021). Quality assessment in systematic literature reviews: A software engineering perspective. *Information and Software Technology*, *130*, 1–24. https://doi.org/10.1016/j.infsof.2020.106397

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Ortega-Ochoa, E., Arguedas, M., & Daradoumis, T. (2024). Empathic pedagogical conversational agents: A systematic literature review. *British Journal of Educational Technology*, *55*, 886–909. https://doi.org/10.1111/bjet.13413