Evolution and trends in intelligent tutoring systems research: a multidisciplinary and scientometric view

Lu Guo^{1,2,4} · Dong Wang¹ · Fei Gu¹ · Yazheng Li³ · Yezhu Wang¹ · Rongting Zhou^{1,4}

Received: 11 November 2020 / Revised: 8 April 2021 / Accepted: 9 April 2021 / Published online: 4 May 2021
© Education Research Institute, Seoul National University, Seoul, Korea 2021

Abstract

Intelligent tutoring systems (ITSs) are a promising integrated educational tool for customizing formal education using intelligent instruction or feedback. In recent decades, ITSs have transformed teaching and learning and associated research. This study examined the evolution and future trends of ITS research with scientometric methods. First, a dataset comprising 1173 relevant publications was compiled from the Web of Science Core Collection databases (including the Science Citation Index Expanded and the Social Science Citation Index). Then, the publication distributions by time, author, institution, country/region, and knowledge sources were analyzed to reveal the multidisciplinary integration paths. Dataset co-occurrence and co-citation analyses were conducted to identify the most popular research issues, the research chronology, and the emerging trends. It was found that: (a) ITS research has been growing in recent years. According to the Price literature exponential growth curve, this field is still in its initial stage while has high potential; (b) computer science, education, psychology, and engineering were the main ITS research knowledge sources, with ITS social science publications since 2007 being higher than ITS natural sciences publications; (c) interactive learning environments, student modeling, teaching/learning strategies, and machine learning have been the most popular research foci; and (d) the Coh-Metrix, problem-centered instruction, and STEM are the current research trends.

Keywords Intelligent tutoring systems · Scientometric analysis · Multidiscipline · Evolution · Trends

Introduction

In recent years, artificial intelligence (AI) has been widely applied in many areas including education (Luckin et al., 2016), with commensurate increases being seen in AI education (AIED) research and applications. AIED adaptive

learning and evaluation applications are now being used to improve educational effectiveness and efficiency (Chassignol et al., 2018; Kurshan, 2016), evaluate teaching effect, adjust teaching and problem-solving strategies in real-time (Shute & Psotka, 1996), and provide a better understanding of student knowledge acquisition (VanLehn et al., 2007; Beal et al., 2010).

ITS research is multidisciplinary involving AI, pedagogy, psychology, and other related disciplines (e.g., Craig et al., 2004; Graesser et al., 2012; Hu & Cooper, 2014; Luckin et al., 2016). Sleeman and Brown's (1982) early book *Intelligent Tutoring Systems* brought together different research fields, with the contributors coming from AI, cognitive science, and education fields (Luckin et al., 2016). Luckin et al. (2016) defined ITSs as computerized learning environments that incorporated computational models from the cognitive sciences, learning sciences, computational linguistics, AI, mathematics, and other fields. Graesser, Conley, et al. (2011), Graesser, Mcnamara, et al. (2011) commented that ITS often incorporated pedagogical, psychological, and other cognitive learning theories into computational models,

Lu Guo and Dong Wang have contributed equally in this work.

✉ Dong Wang
wang0817@mail.ustc.edu.cn

✉ Rongting Zhou
rongting@ustc.edu.cn

¹ School of Humanities and Social Sciences, University of Science and Technology of China, Hefei 230026, China

² School of Humanities, Hefei Normal University, Hefei 230061, China

³ School of Journalism and Communication, Anhui University, Hefei 230601, China

⁴ Key Laboratory of Immersive Media Technology (Wanxin Media), Ministry of Culture, Hefei 230051, China

Cristina Conati (2002) noted that ITS research focused on advances in AI, cognitive science, and education to improve computer-supported education, and Ahuja and Sille (2014) commented that ITS research was an intersection of computer science, cognitive psychology, and educational research.

As early as the 1970s, AI systems were being used to provide individual, adaptive instruction (Conati et al., 2002a, 2002b; Luckin et al., 2016). The first ITS system, which was developed by Carbonell in 1970, was SCHOLAR (Dargue & Biddle, 2014), which was followed a few years later by several influential ITSs, such as the BIP designed by Stanford University in 1977 (Wescourt et al., 1977), WUMPUS developed by MIT in 1977 (Xu et al., 2009), SOPHIE (Sleeman & Brown, 1982), DEBUGGY (Sleeman & Brown, 1982), and AutoTutor (Graesser et al., 2005). The teaching effectiveness of these ITSs has been found to nearly parallel real-life teachers (VanLehn, 2011) and enhance the performances of both teachers and students (Spector et al., 2014). Therefore, ITSs are being regarded as the propeller for the future of human education (Luckin et al., 2016).

Conceptually, there is no clear boundary between ITSs and related concepts such as computer-assisted learning (CAL), computer-assisted training (CAT), and computer-assisted instruction (CAI) (Sleeman & Brown, 1982; VanLehn, 2006; Anderson et al., 1990). However, Conati (2009) claimed that the key difference between ITSs and CAI was that the solutions provided by ITSs were generated in real-time from student input, rather than having to be predefined, and Wenger (1987) and VanLehn (2011) commented that while CAI, CAL, and CAT were based on responses, ITSs were based on steps. In this study, CAI-related concepts have been excluded.

ITSs research has been multidisciplinary, with the key research associations being computer science and computer technologies and the core of the research being the development of student models (Desmarais & Baker, 2012) that digitize student abilities and allow them to access personalized instruction (Conati & Kardan, 2013) matched to their aptitude (Graesser et al., 2012; Hu & Cooper, 2014; Vandewaetere et al., 2011). The benefits of these ITS student models and other components are that they promote personalized learning, provide real-time learning analysis, use self-adaptive content, and designate targeted practice (Liu, 2003). Corbett and Anderson (1994) first proposed a knowledge tracing (KT) model based on a hidden Markov model to identify changing cognitive states during knowledge acquisition by analyzing student data and predicting future performances.

Another crucial research field has been ITS applications (D'Mello et al., 2007) such as AutoTutor, which has also included comparative studies that compared ITS learning effects with other forms (VanLehn et al., 2007) to assess

effectiveness. Some studies found AutoTutor to be as effective as a human tutor in computer-mediated conversations (Graesser et al., 2003; Person et al., 2001), and others identified the specific factors related to the gains in deep-level comprehension (Graesser et al., 2009; Baker et al., 2010). These early ITS research studies attracted greater research interest and the development of other ITS applications, such as Coh-Metrix and Cognitive Tutor (Graesser et al., 2014; Graesser, Conley, et al., 2011; Graesser, Mcnamara, et al., 2011; McNamara et al., 2010, 2014; Pane et al., 2014).

Scholars in psychology provided theoretical tools and was of great significance to this kind of research (Arroyo et al., 2009; D'Mello et al., 2007). For example, Kort et al. (2002) proposed a comprehensive four-quadrant model that explicitly linked learning and affective states), Anderson proposed an adaptive control of thought (ACT) cognitive theory (Anderson, 1980, 1983), which became the theoretical basis for the popular Cognitive Tutor ITS system, and others have used ITSs to identify the different influences that different affective/cognitive states have on learning effects. (Craig et al., 2004; Koedinger & Corbett, 2006; Woolf, 2009; Arroyo et al., 2009; Lester et al., 2013; D'Mello & Graesser, 2011).

Pedagogical researchers have focused on designing teaching strategies to achieve better teaching effects (Kolodner, 2002; Luckin et al., 2016; VanLehn, 2006). For example, Graesser et al. (2005) reviewed the pedagogical strategies that embedded constructivist approaches in ITS instruction, finding that the learning effect was negatively correlated with boredom and positively correlated with flow, and also explored the relationships between the emotional state and the learning process. Other pedagogical aspects have been examined to improve the effectiveness and efficiency of education, such as game-based learning strategies (Lester et al., 2013; Tsay et al., 2018; Santhanam et al., 2016), collaborative learning environments (Chi et al., 2001), and intelligent narrative technologies (McCoy et al., 2011; Yu & Riedl, 2012).

Given all this interest over several decades, there have been a number of previous ITS research reviews, such as systematic reviews focused on ITS composition, current research foci, and current trends (Almasri et al., 2019; Akbulut & Cardak, 2012; Baker et al., 2008; Schmidhuber, 2015), and reviews focused on technology/evaluation methods in different ITSs (Desmarais & Baker, 2012; Elham et al., 2018; Conati, 2002a, 2002b). In a recent review, Elham et al. (2018) concluded that the most frequent AI techniques applied to ITSs were action condition rule-based, Bayesian networks, and data mining. Other reviews have given suggestions on specific teaching mechanisms, such as strategic decision making based on student emotions (Sharma et al., 2014; Conati, 2002). The most promising ITS research trends have been identified as

portable devices (Elham et al., 2018; Ahuja & Sille, 2014) and collaborative learning (Isotami & Misogici, 2008), with others redefining the purpose of using ITSs as learning devices that are unable to be effective without human guidance rather than systems to improve effectiveness and achievement rates (VanLehn, 2011; Woolf, Lane, et al., 2013; Woolf, Chad Lane, et al., 2013). Another common ITS research review type has been meta-analyses based on quantitative systematic reviews, most of which have compared the learning effects of different ITS systems, the teaching methods, or the learning conditions (VanLehn, 2011; Graham et al., 2015; Steenbergen-Hu & Cooper, 2014; Ma et al., 2014; Kulik, 2015).

The review conducted in this paper employed a scientometric method, which uses bibliometrics such as citation analyses to evaluate the scientific research activities that guide science-policies (Egghe, 2005). Nalimov and Mulchenko (1971) coined the term “scientometrics” in the 1960s to describe the growth, structure, interrelationships, and productivity in scientific research: (Hood & Wilson, 2001), after which scientometrics was primarily employed to analyze research literature based on the attributes in the research itself, such as the number of publications, keywords, or other dynamic indicators such as citation information. Compared with meta-analyses or systematic reviews, which usually require detailed coding or the weighting of research content based on human judgment, the scientometric method automatically calculates and demonstrates the information based on the publication attributes, that is, it is an analysis method based on data and algorithms.

Scientometrics can provide a quantitative and systematic review of all aspects of ITS research, such as the publication and disciplinary states, the particular research issues, the intellectual structures, and the emerging trends. While previous research has discussed the multidisciplinary composition of ITS research, it has not provided a discipline construction path (Craig et al., 2004; Koedinger et al., 2012; Woolf, 2009). Therefore, this review will reveal the path of the multidisciplinary integration of ITSs and the proportions of the composition. The proportions involved in the composition of ITSs researches and the path of its formation are essential to understand the characteristics of ITSs as a subdiscipline as well as to identify the discipline formation stage. Only by analyzing the characteristics and composition of the subdiscipline, the evolution of the knowledge sources, and the intellectual base can researchers be better guided concerning the research methods or skills they need to master. Further, the illumination of the current stage of the subdiscipline can help researchers identify more promising research directions and adopt more suitable strategies.

Therefore, based on the following research questions (RQs), this research aimed to reveal the history, current

status, and trends in ITS research from a scientometric and multidisciplinary perspective:

RQ1: What is the current status of ITS research? What are the main contributing countries/regions and the major journals, authors, and institutions?

RQ2: What are the ITS knowledge sources and how have they developed?

RQ3: What have been the most popular ITS research foci?

RQ4: What are the chronological ITS research stages and the intellectual bases in each of these stages?

RQ5: What are the emerging ITS research trends?

Data and methods

Data collection

The relevant research article data for this study were extracted from the Science Citation Index Expanded (SCIE) and the Social Science Citation Index (SSCI) in the Web of Science (WoS) Core Collection database from 1963 to 2020. Due to deficiencies in the originality, completeness, and impact of review articles, conference papers, proceedings, and other document types, only research articles were included in this study. Review articles were excluded because they were not original research and their inclusion would possibly lead to over-citation and overrepresentation, especially for highly cited papers (Miranda & Garcia-Carpintero, 2018), which could have affected the veracity of the results.

The 1173 bibliographic records (research articles) and 12,992 associated references were collected on May 21, 2020, using the advanced search service offered by the WoS Core Collection. A query formula that included field tags, Boolean operators, parentheses, and query sets was created to retrieve the desired literature, as follows: TS = (“intelligent tutoring systems” OR “Intelligent Computer-Aided Instruction” OR “Intelligent Computer-Assisted Instruction” OR “Artificial Intelligence in Education” OR “adaptive educational system” OR “adaptive learning systems” OR “constraint-based tutors” OR “Cognitive Tutor” OR “AutoTutor” OR “SQL-tutor” OR “assistments” OR “elm-art” OR “iweaver” OR “DeepTutor” OR “Coh-Metrix” OR “Electronix Tutor” OR “Student modeling” OR “Knowledge Tracing”). This retrieval formula included concepts, theories and methods, and applications to ensure precision in the ITS research retrieval results.

Data analysis

After the data collection, several descriptive statistical analyses were performed to identify the WoS categories,

the specific country/region, journal, author, institutional, research area publication numbers, and the ITS research field citation distribution. To explore the information behind the publications, a series of co-occurrence analyses (co-word and co-citations) were also conducted.

Co-occurrence analyses or co-word analyses were conducted on the keywords. The keyword network in scientific literature can be used to explore the correlations between different research studies to reveal the potential research issues and the intellectual structure in a certain research field (Chen et al., 2019). This method has been widely used in bibliometric studies (Wu et al., 2016) and has proven effective (Khasseh et al., 2017).

Then, a co-citation analysis of the articles was performed. Two articles that are cited in the same article are believed to have a certain relevance, that is, a co-citation relationship, as these co-cited articles are the upstream knowledge sources. Co-citation analyses are not only capable of presenting the intellectual evolutionary process in the ITS field, but can also reveal the milestone literature that promote innovation (Hou et al., 2018).

Visualization and tools

The scientometric research tools, CiteSpace and VOSviewer, were used to visualize the co-word and co-citation networks to elucidate the deeper structures in these networks.

Developed by Chen (2006), CiteSpace has been widely used for bibliometric analyses and visualizations. In addition to co-occurrence analysis and data visualization, CiteSpace also generates cluster analyses and cluster-labeling using algorithms such as the Log-Likelihood Ratio (LLR) to extract the themes in each cluster and also uses Burst detection to identify emergent research-front concepts. VOSviewer, which was developed by Van Eck and Waltman (2009), was also employed to construct and view the bibliometric maps.

This paper harnessed the strengths of each of the above-mentioned tools to enhance the presentation of the study results. Table 1 shows the corresponding relationships

between the research questions, data analysis methods, tools, and research content.

Results and discussion

Publication analysis

RQ 1: Three-stage publication growth

Figure 1 shows the number of publications in each year. A total of 1173 articles (publications) were collected. The earliest paper appeared in 1963 (Cavanagh, 1963), but besides this paper, few relevant papers were published in the 1960s and 1970s. However, the general ITS technological development during the early years was driven by AI technology.

More ITS-related publications appeared in the 1980s and since then, the number of publications has been keeping fluctuant growth. The overall distribution of ITS-related publications had a three-stage growth. During the first stage (1985–1998), the ITS research studies jumped from 7 in 1989 to 32 in 1998, with a mean of 14. In the second stage (1999–2006), after a brief decline in 1999, the number of publications increased, with a mean of 32.6, with a peak appearing in 2006. However, it was then followed by a sharp decline in 2007, but rebounded in 2008. The third stage was from 2007 to 2019, during which time, the mean was 52.4, and reached an all-time peak of 77 in 2013. In general, the number of publications has had a fluctuating growing trend, with the average number of publications in the three stages being respectively 14, 32.6, and 52.4, from which the three-stage growth pattern can be clearly seen.

There were several noteworthy AI technology events that increased public interest in AI technology and led to technological breakthroughs. As the public attention began to attract research interest, there was a significant increase in the number of publications.

Lighthill's report (1973) revealed that there was a gap between AI technological expectations and reality, which resulted in a reduction in funding in some countries for the

Table 1 Relationship between research questions, methodology, and analysis tools

RQ	Methodology	Analysis tools	Justification
RQ 1	Descriptive statistic	Excel 2016	It presents the distribution of publications, citations, countries/regions, journals, authors, institutions.
RQ 2	Descriptive statistic	Excel 2016	It shows the constitution of ITSs Research Areas which are assigned by WoS tag.
	Category timezone	CiteSpace 5.5.R2	It shows the development of ITSs Categories that are assigned by WoS tag.
RQ 3	Co-word analysis	VOSviewer 1.6.15	The keyword co-occurrence clusters are made to present the research issues.
RQ 4	Co-cited reference	CiteSpace 5.5.R2	The co-cited references represent the intellectual base, which can be used to illustrate the knowledge structure of ITSs.
RQ 5	Burst detection	CiteSpace 5.5.R2	Extract the latest burst co-cited references to depict the trends in ITSs research by the Burst detecting function.

Fig. 1 ITSs topic papers collected from SCIE and SSCI, 1963–2020. *Note.* The Top Edge of the Yellow Rectangle represents the mean number of publications at each stage

first ebb since the Dartmouth Artificial Intelligence Conference in 1956. In 1986, the development of the back propagation (BP) neural network algorithm (Rumelhart et al., 1986) incited new interest in ITS research. In 1997, Deep Blue defeated the world chess champion, which was a milestone in AI development and awakened public and research interest in related ITS applications, with the number of publications climbing to 37 in 1998. However, a year later, only 11 papers were published as chess robots were not seen as directly relevant to ITSs. In 2006, Hinton's deep belief network (Hinton et al., 2006), which was a breakthrough in neural network algorithms, spurred renewed interest, with the number of publications increasing to 59, and in 2013, the deep learning algorithm that finally achieved speech and visual recognition success, resulted in the era of perceptual intelligence (Li, 2018), which resulted in publications reaching an all-time peak.

These landmark events were respective ITS publication triggers, which little by little led to the fluctuating growth in the number of publications over the years. The relationship between AI technology and ITSs was due to the obvious spillover effect as cutting-edge AI theory and technological research started to lead to ITS education application developments. As AI research increased, interest in ITSs increased, with each research field having an influence on the other as reflected in the fluctuating number of publications. From the scientometric view, the stepped and fluctuating growth in ITS research (Fig. 1) did not conform to the Price literature exponential growth curve (Tague et al.,

1981), which demonstrated that ITS research was still in its initial discipline formation stage.

Since the neural network algorithmic breakthroughs in 2006, algorithms such as Deep Learning, CNN (Convolutional Neural Networks), and GAN (Generative Adversarial Networks) have gradually become the most exciting AI areas. In recent years, one of the deep learning algorithms, CNN, which specialized in recognizing images, facial expressions, voice, and even text, has begun to be more widely applied to education through ITS research optimized model designs and parameter values (Liu et al., 2018; Hu et al., 2020). Another deep learning algorithm, GAN, which has also received significant attention, has seldom been used in education (Chen et al., 2020), but is suitable for automatically generating graphics and audio images and could be a useful future ITS application.

RQ 1: Publication numbers in SCIE or SSCI

ITS research combines achievements from both the social and natural sciences. Figure 2 compares the number of publications in the SCIE and SSCI databases from 1963 to 2020, which are respectively related to natural and social science publications, with 330 publications included by both indices. However, this section focuses on the relationships between the social and natural science ITS publications.

The SSCI publications in this field first appeared in the 1960s, which was far earlier in the SCIE papers (1986). This revealed that the possibility of ITSs was first examined in the

Fig. 2 SCIE and SSCI ITS-related publication numbers (1963–2020)

social sciences. Before 2007, the number of SCIE publications exceeded the number of SSCI publications, but after 2007, the two domains had commensurate growth until 2010, when the number of SSCI publications surpassed the number of SCIE publications.

The first surge in ITS-related SCIE publications in 1992 was possibly because of engineering research. The second SCIE surge was in 1997, which was the same year that the Deep Blue chess AI defeated the human chess master for the first time, which initiated a global upsurge in AI research. From 1992 to 2006, the number of natural science publications was significantly higher than in the social sciences, but after 2007, the number of social science publications exceeded those in the natural sciences.

RQ 1: Exponential increase in citation trends

The annual number of ITS citations had an exponential growth trend (see Fig. 3), increasing from 4 in 1986 to 2714 in 2019, with this growth trend being precisely fitted using the software Origin 2019 Pro with $R^2 = 0.99$. The rapid growth of the fitting curve started from around 1990, which marked the beginning of a significant increase in ITS citations.

RQ 1: Top productive countries/regions, authors, journals and institutions

Table 2 lists the top five dominant ITS research countries/regions and authors. The ITS research geographic distribution can help researchers identify the countries/regions on the cutting edge of ITSSs. The United States ranked first with 480 publications, or about 40.9% of the total, with the most

Fig. 3 Growth of citations

Table 2 Top productive countries/regions and authors

Rank	Country/Region (NP, Prop.)	Authors (Country, NP, H-index)
1	USA (480, 40.9%)	A.C. Graesser (USA, 52, 43)
2	CANADA (85, 7.2%)	D.S. McNamara (USA, 34, 40)
3	SPAIN (83, 7.1%)	K.R. Koedinger (USA, 27, 32)
4	TAIWAN (76, 6.5%)	V. Aleven (USA, 24, 25)
5	ENGLAND (74, 6.4%)	A. Mitrovic (UK, 24, 16)

productive writer being Graesser, a professor at the University of Memphis specializing in psychology, discourse, language, AI, and education. The top 4 productive authors were all from the USA.

Table 3 shows the top five journals that published ITS-related research and the institutions with the highest number of ITS publications. *Computers & Education*, which is an interdisciplinary journal for different types of technically-based education studies, published the highest number of ITS papers. Most of the top five journals had research studies focused on computer science, education, engineering, psychology, cybernetics, and other disciplines, which reflected the multidisciplinary features of the ITS domain. The University of Memphis published the most ITS studies.

Multidisciplinary characteristics based on WoS category

The journal categories for the ITS publications were generated from the WoS classification tags “Research Areas” and “WoS Categories,” with the former giving the general journal classification, and the latter giving more detailed information that could be detected by the CiteSpace category function.

RQ 2: Distribution of research areas and the Timezone view of categories

In the top ten research areas based on the ITSs WoS classifications by “Research Areas” tags, four areas significantly prevailed: “COMPUTER SCIENCE” (688 publications); “EDUCATION & EDUCATIONAL RESEARCH” (453 publications); “PSYCHOLOGY” (183 publications); and “ENGINEERING (159 publications).” The number of publications in the computer science domain was much higher than in the other research areas, and these four research areas appeared to be the fundamental ITS research branches.

To dynamically illustrate the multidisciplinary ITS research integration paths, CiteSpace software was applied to develop a time zone view (see Fig. 4) of the WoS Categories in this field. In Fig. 4, the size of each circle is proportional to the number of publications in the category, and the position of the node centers indicate the year in which the initial publication in each category appeared.

As the journal that included the earliest publication in 1963 (Cavanagh, 1963) had incomplete information, it was filtered out from the category time zone (Fig. 4). In 1964, the Education & Educational Research and Education, Special research fields first appeared, along with a publication by Cavanagh (1964). In 1986, Psychology and Computer Science appeared, followed by Computer Science and Information Systems (1988), Computer Science and Interdisciplinary Applications (1989), Computer Science and Computer Science Theory & Methods (1989), Computer Science Artificial Intelligence (1990), Computer Science and Cybernetics (1994), and Computer Science and Software Engineering (1994), with these categories forming the largest community. Since then, research related to computer science has been the most significant knowledge source for ITS research. In 1991, engineering research appeared, including Engineering and Engineering, Electrical & Electronic, and in 2008, ITS-related linguistics research emerged.

As shown in Fig. 4, the ideas for ITSs originated from the education discipline. In the 1960s, pioneering education scholars introduced a rudimentary ITS application. Due to the scarcity in basic research achievements and computer hardware, ITSs research came to a standstill for decades. However, from the middle of the 1980s, a group of scholars introduced computer technology theories and methods into ITS research, and in 1986, the introduction of psychological theories and methods began to contribute significantly to the study of evaluation. In 1991, ITSs appeared in the “Engineering” and “Electrical & Electronic” fields, which were mainly focused on software engineering or automated/information systems. For example, an article published in Information and Software Technology entitled “SQL Tutor, a co-operative ITS with repository support” introduced the system architecture and concluded that repository technology was a technical solution that supported multi-user co-operation and collaboration (Wang, 1997).

The most recent subject, CHEMISTRY, TELECOMMUNICATIONS appeared in 2019, followed by Psychology Mathematical (2016), Psychology Multidisciplinary (2013), and Linguistics/Language (2008). The Category

Table 3 Top productive journals and institutions

Rank	Journals (IF 2018, NP, Prop.)	Institutions (NP)
1	Computers and Education (5.6, 119, 10.1%)	University of Memphis (74)
2	User Modeling and User Adapted Interaction (3.4, 52, 4.4%)	Carnegie Mellon University (69)
3	IEEE Transactions on Learning Technologies (2.3, 41, 3.5%)	Pennsylvania Commonwealth System of Higher Education (49)
4	Expert Systems with Applications (4.3, 35, 3%)	Arizona State University (38)
5	Computers in Human Behavior (4.3, 32, 2.7%)	University of Pittsburgh (37)

NP represents the Number of Publications, Prop. represents the proportion of the retrieved 1173 articles, IF represents the Impact Factor in 2018

Fig. 4 Category time zone of ITSs research

Timezone view revealed the path of multidisciplinary integration of ITS.

Computer science was the main ITS function background. While the theoretical basis for ITSs was education and psychology, the integration needed to be realized through technology. Therefore, the related computer science studies were focused on specialized computer science skills and the incorporation of educational or psychological frameworks. For example, Chang et al. (2020) proposed methods to maintain the benefits of semantic web-based approaches when representing ITS pedagogical rules, and Liu et al. (2018) incorporated Cognitive Diagnosis models, which were an evaluation system based on cognitive psychology, statistics, and computer science, into an ITS to model to analyze student answer data, which introduced a correlation between test questions and knowledge structures to diagnose student cognitive states and quantitatively investigate student differences and cognitive levels.

Algorithms need to be guided by educational and psychological theories, which then need to be properly embedded in the designed teaching practice systems and programs. It is also of great significance to have experts in the fields of education and psychology evaluate methods using various ITS applications, propose suggestions for system improvement during the implementation, and develop theories applicable to the educational context.

Some scholars have commented that to adapt to different teaching situations, pedagogical and domain models based on embedded pedagogical or psychological theories were essential for ITSs (Luckin, 2016; Sharma et al., 2014). For example, Scaffolding theory and Socratic Questioning have been used to design a pedagogical ITS model (Azevedo & Hadwin, 2005; Cotton, 2001; VanLehn, 2011). From an application perspective, ITSs can provide adaptive and personalized teaching strategies for different situations. Graesser et al. (2011) concluded that the ITSs environment was well suited to the acquisition of pedagogical strategies and deeper knowledge in the cognitive spectrum), and Anderson et al. (1995) examined the development of individual instruction, arguing that the best tutorial strategy was to provide immediate feedback in short and directed error messages).

ITSs have also been found to be advantageous in tracking psychological attributes, such as emotional/cognitive recognition, and in developing diagnostic information models (Chang et al., 2020; Conati & Maclaren, 2009; Craig et al., 2004) and computer perception devices that automatically monitor student emotions while they are learning (Aleven et al., 2006; Graesser & d'mello, 2012).

In terms of theoretical constructs in psychology, Matz (1981) developed one of the first detailed psychological models to explore why students had certain misconceptions while using ITSs, which provided a cornerstone for building

Research issues

Because keywords are refined research content, popular ITS research foci can be identified using keyword co-occurrence analysis (Xie, 2015), which indicates that the keyword appears in the same article. When two keywords appear in the keyword list for the same article, a co-occurrence relationship is established. Excluding the phrase “Intelligent tutoring systems”, the top 15 keywords in descending order of co-occurrence frequency were as follows: interactive learning environments (69); student modeling (51); teaching/learning strategies (35); machine learning (24); human–computer interface (23); evaluation methodologies (19); e-learning (19); architecture for educational technology systems (18); Bayesian networks (17); adaptive learning (16); learning (16); artificial intelligence (15); Coh-metrix (15); multimedia/hypermedia systems (14); improving classroom teaching (14); and computational linguistics (14). Each keyword here is derived from research articles in the ITSs field collected for this study.

The red cluster was the largest with 32 items and served as the bridge to the other three clusters. The dominant topic in the red cluster was emotion recognition due to the keywords including “learning,” “affective computing,” “affect, emotions,” and “cognition.” The green cluster (26 items) contained the nodes such as “student modeling,” “machine learning,” “Bayesian networks,” “educational data mining,” and “fuzzy logics,” which implied that the core cluster topic was computer technologies. In the blue cluster (21 items), the most frequent co-occurrence keyword was “interactive

Fig. 5 Keyword co-occurrence networks

learning environments,” with the other keywords in this cluster being focused on teaching/learning strategies, educational technology architectures, human–computer interfaces, and pedagogical issues as well as “post-secondary education” and “secondary education,” indicating that the research in the two themes were tightly associated with the ITS learning environments and learning strategies. The yellow cluster (14 items) contained the nodes, “Coh-Metrix,” “readability,” “coherence,” and “cohesion” and so on. Coh-Metrix is an online text analysis tool designed and developed by the University of Memphis that measures readability, vocabulary, syntax, concreteness, cohesion, and the story hood of a text.

The red keyword co-occurrence network cluster indicated that the ITS cognitive, emotional, and educational issues were closely related, that is, cognitive psychology and ITS education appeared to be highly integrated. However, while the analysis did not identify a keyword for an exclusive ITS research theory, it was inferred that there were few guiding theories. Therefore, it appears that there is an urgent task for researchers to develop targeted pedagogical and psychological theories for the ITS context.

Most of the keywords in the green cluster were general rather than specific technical types, such as “machine learning,” “Bayesian networks,” and “fuzzy logic,” which indicated that because of the diverse application types and technologies in ITSs, the research was not specifically focused on a specific technology. So the keywords were only reflected the most common technology types. Moreover, the latest AI technologies that were leaps and bounds from artificial neural network algorithm since 2006 were not reflected, which was possibly because of the publication research lag. However, it did indicate that greater research attention is necessary on emerging algorithms and technologies such as GAN.

Intellectual structure evolution

The intellectual structure of a given research field can be illuminated using a co-citation reference analysis (Chen et al., 2010), which highlights the key fields in the intellectual base and predicts the intellectual structure trends (Chen et al., 2008; Madani & Weber, 2016). When two publications are cited by another publication, these two publications are seen to be interrelated (Griffith et al., 1974; Small & Griffith, 1974). The collection of references cited by a research community, therefore, comprises the intellectual base of the research frontier (Braam et al., 1991). Co-citation analyses identify the consistency of the concepts and references and their associations with the research field (Anwar et al., 2019).

In this section, the intellectual structure evolution is demonstrated through a visualization of the intellectual base and the research fronts (Chen & Guan, 2011). The co-citation

analysis was based on the 12,992 references in the 1173 sample publications.

RQ 4: Chronological research stages

The co-citation references and their relationships are the nodes and links in the co-citation network, which are clustered based on the co-citation associations (Madani & Weber, 2016). In CiteSpace, the co-citation network clusters are labeled by the Log-Likelihood Ratio algorithm (Chen et al., 2010), which reveals the main research specialties in the nodes in each cluster. (Chen, 2017; Hu, 2017). Each cluster has a specific ID, such as #0, #1, as shown in Fig. 6. The ID numbers are in descending order based on the size of the cluster (Li & Chen, 2016), and the color bar below Fig. 6 changes from gray to red to represent the transition from 1963 to 2020.

The silhouette score in Table 4 is an indicator of the cluster homogeneity, that is, when the silhouette score is greater than 0.7, the clustering results have high reliability (Li & Chen, 2016). *The Mean Year* is the average publishing year of the literature in the cluster, which is used to detect the ITS intellectual base evolution. The top terms are detected by the Log-likelihood algorithm from the keywords in the co-cited references. The top terms in Table 4 are the first two terms in each cluster that were statistically significant.

From the cluster evolution shown in Fig. 6, the co-cited reference clusters were divided into three stages in Table 4, which are shown in different background colors.

Clusters #7, #11, #9 and #4 constituted the early stage (mean years 1986–1992). Using the LLR tag semantic analysis, it can be seen that the early research was mainly focused on the ITS applications associated with knowledge representation and CAI, such as adaptive systems and business management gaming simulations, and the development of cutting-edge computer technologies, such as unsupervised learning and neural networks. This early stage research indicated that ITS research originated from an education base and was established through the associations of education and computer technologies, with the core themes being related to how the education and computing technologies could be combined.

Clusters #5, #8, #4, #6, #23, #13, #10, #18, #16 and #0 constituted the second stage (mean years 1997–2006), with the research themes being an extension of the earlier stage. Due to the spillover effect of the continuing technological progress, many ITS applications such as AutoTutor and Cognitive tutor began to emerge. Consequently, the research directions began to focus on the integration of pedagogy, cognitive psychology, and linguistics theories into ITS applications (Anderson et al., 1990; Johnson & Richel, 2000). The 2006 publication were related to AI technology breakthroughs (Hinton et al., 2006) and the application of deep

Fig. 6 Co-cited reference clusters with cluster ID

Table 4 Information of the co-cited references clusters

Cluster ID	Mean year	Cluster size	Silhouette	Top terms (LLR, p-level)
7	1986	44	0.98	Adaptive systems (13.33, 0.001); Individual differences (6.73, 0.01)
11	1987	27	0.996	Knowledge representation (7.98, 0.005); Computer assisted instruction (6.65, 0.01)
9	1992	33	0.955	Business management gaming simulation (6.82, 0.01); Unsupervised learning (6.82, 0.01)
20	1992	10	0.984	Neural networks (computers) (7.84, 0.01); Decision support techniques (7.84, 0.01)
5	1997	50	0.886	Probabilistic models (6.02, 0.05); Adaptive testing (6.02, 0.05)
8	1998	40	0.933	Computer-mediated communication (12.39, 0.01); RDF (12.39, 0.01)
4	2000	17	0.983	STEM learning (8.54, 0.005); Multimedia/Hypermedia system (8.54, 0.005)
6	2002	10	0.903	Computer-supported collaborative learning (11.34, 0.001); Learning styles (10.58, 0.005)
23	2004	31	0.709	Semantic web-based educational systems (7.11, 0.01); Geometric reasoning (7.11, 0.01)
13	2005	23	0.983	Latent semantic analysis (13.73, 0.001); Coherence (6.82, 0.01)
10	2005	33	0.925	Cognitive tutors (7.36, 0.01); Metacognition (5.76, 0.05)
18	2005	11	0.996	Meta-cognitive skills (9.3, 0.005); Scientific inquiry learning (9.3, 0.005)
16	2006	14	0.963	Data mining (6.35, 0.05); Sequential analysis (6.35, 0.05)
0	2006	114	0.794	Human-computer interaction (13.52, 0.001); Autotutor (13.52, 0.001)
3	2009	57	0.785	Model construction (9.71, 0.005); Fuzzy-trace theory (6.08, 0.05)
1	2011	62	0.959	Computational linguistics (13.57, 0.001); Writing assessment (12.81, 0.001)
14	2012	22	0.949	Casual path analysis (10.87, 0.001); Student modeling (5.93, 0.05)
15	2012	22	0.971	Pedagogical agent (11.14, 0.001); Out-of-system transfer of learning (5.55, 0.05)
2	2012	60	0.898	Problem-centered instruction (14.72, 0.001); STEM (14.72, 0.001)

Note. Mean Year 1986–1992 represent the first stage, Mean Year 1997–2006 represents the second stage, and Mean Year 2011–2012 represents the third stage. *LLR* Log-likelihood Ratio

learning technologies to ITSs (Conati, 2009; Desmarais & Naceur, 2013; Koedinger et al., 2012). Research also strengthened and deepened in various subfields as shown in the several co-citation reference cluster branches, such as #0 human–computer interactions, #5 probabilistic models, #8 computer-mediated communication, #10 Cognitive tutors, #18 meta-cognitive skills, #23 semantic web-based educational systems, #13 latent semantic analyses, #4 STEM learning, #16 data mining, and #6 computer-supported collaborative learning. Of these, the research on human–computer interactions was the largest research cluster in this stage.

The appearance of these multiple research branches in the second stage was directly related to AI technological breakthroughs, which opened up the field to different types of ITSs. For example, the progress made in the development of probability models improved the construction of student models, and the multidisciplinary pedagogy, cognitive psychology, and linguistics theory developments deepened the ITS application research possibilities.

Two research branches appeared as the intellectual bases in the third stage (# 1, # 3, # 14, # 15, # 2). Cluster #1 was associated with developments in computational linguistics and was an extension #6 and #23 and semantic analysis. The references in cluster #1 were mainly associated with ITS text analysis applications such as Coh-Metrix. The second branch (# 3, # 14, # 15, # 2) was an extension associated with the combination of computer technology and pedagogy in ITSs, such as problem-centered instruction, and STEM. Some of the themes that appeared in the clusters #4 and #5 reappeared in clusters #2, #3, and #14, which suggested that the STEM learning and model constructions were common ITS research concerns. The two research branches in the third stage, therefore, appeared to be related to computer technology spillover effects, as they represented progress in areas such as the associations between NLP (Natural Language Processing) and social science and ITS text analysis applications such as Coh-Metrix; that is, the research in this stage deepened the research directions identified in the previous stages.

The scientometrics analyses of the ITS research revealed the multidisciplinary stage of discipline formation, as highlighted in Shneider's theory (2009). Shneider proposed that the discipline formation may experience four stages, the initial conceptualization stage, the multidisciplinary stage, the expansion stage, and the final stage of decay (Shneider, 2009). The researchers were applying methods from other disciplines to deepen the ITS application-oriented research. Several studies (Choi & Clin, 2006; Núñez et al., 2019); Garnder (1987) claimed that multidisciplinary science was a weaker version of interdisciplinary research, which creates its own theoretical, conceptual, and methodological identity (Núñez et al., 2019). However, it could also be argued

that ITS research is currently multidisciplinary as specific ITS communication systems or specific theories have not yet been developed. Overall, the main characteristics of the studies in this period were originality and creativity, which tended to indicate that to effectively extend into new research areas, researchers needed to practice high-risk endurance when choosing their tasks (Shneider, 2009).

RQ 4: Significant references in the co-citation networks

Frequently co-cited references imply advanced ideas and developments in a given research field (Anwar et al., 2019). Table 5 shows the top 10 most frequently co-cited references and gives a brief description of the articles. Articles with high betweenness centrality scores generally indicate (Chen & Guan, 2011) a fundamental transition in the research knowledge domain paradigm and the significant influence these references have on the co-citation structure (Li & Chen, 2016). The betweenness centrality in CiteSpace is algorithmically identified and represented using magenta circles, which allows researchers to understand the changes in a domain's knowledge structures over time (Chen, 2005).

The article with the highest betweenness centrality (centrality = 0.32) was “Cognitive tutors: Lessons learned” (Anderson et al., 1995), which presented several empirical studies on ITS learning effects, the development of procedures when cognitive models were incorporated into tutoring systems, and the development of individual instruction. The article also reviewed the use of KT and Bayesian technologies to assess the probabilities that students had learned the principles in the cognitive model. This research article belonged to cluster #5 (probabilistic models, adaptive testing) at the connection point of #9, #5, and #8 in Fig. 6. Therefore, this study was a notable joint in the intelligence base and transited single chain development to the multiple sub-research fields.

RQ 5: Emerging ITS trends

Emerging research-front concepts were identified using CiteSpace's Burst detection of the article citations (Chen, 2006), and were specifically detected using the Kleinberg algorithm (Feng et al., 2015). Co-cited references with high Burst values indicate that the number of co-citations of an article suddenly increased sharply (Chen, 2012; Hou et al., 2018). The citation reference themes with high Burst values in a specific period, therefore, reflect the research trends in specific domains for the next few years (Hou et al., 2018).

Figure 7 shows the co-cited reference areas with high Burst values, most of which were located between 2019 and the first half of 2020, with the red dot representing those references with high burst values. The image in the larger red circle is an enlarged drawing of the area within the smaller

Table 5 Top 15 co-cited references

No	Type	Authors	Co-citation frequency	Title	Findings
1	Review	K. VanLehn (2011)	54	The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems	Find that the effect size of human tutoring $d = 0.79$. The effect size of intelligent tutoring systems was 0.76, so they were nearly as effective as human tutoring
2	Article	A. C. Graesser, D. S. McNamara, J. M. Kulikowich (2011)	46	Coh-Metrix: Providing Multilevel Analyses of Text Characteristics	Figured out 5 main factors that account for most of the variance in texts across grade levels and text categories: word concreteness, syntactic simplicity, referential cohesion, causal cohesion, and narrativity using Coh-Metrix
3	Book	D. S. McNamara, A. C. Graesser, P. M. McCarthy, & Z. Cai (2014)	32	Cambridge University Press: Automated Evaluation of Text and Discourse with Coh-Metrix	The book explored the development, measures of Coh-Metrix, and the empirical work conducted with Coh-Metrix at first. Then it described how to use Coh-Metrix and how to analyze and interpret the results
4	Article	D. S. McNamara, M. M. Louwerse, P. M. McCarthy, & A. C. Graesser (2010)	31	Coh-Metrix: Capturing Linguistic Features of Cohesion	Find that Coh-Metrix indexes of cohesion (individually and combined) significantly distinguished the high- versus low-cohesion versions of texts. The commonly used readability indexes (e.g., Flesch-Kincaid) were inappropriately distinguished between high and low-cohesion texts
5	Book	E. Wenger (1987)	31	Artificial intelligence and tutoring systems	An introductory book, first elaborated on knowledge communication systems, basic issues, and tutorial dialogues
6	Article	K. VanLehn (2006)	27	The Behavior of Tutoring Systems	The tutoring systems were described as two loops. The outer loop executed once for each task, where a task usually involves solving a complex, multi-step problem. For each step the student took in solving the task, the inner loop gave feedback and prompt at each step. The inner loop also evaluated the students' developing ability and updated the student model. The outer loop could use the model to select the next task suitable
7	Book	T. K. Landauer et al. (2007)	26	Handbook of Latent Semantic Analysis	A handbook that explored crucial aspects (e.g., theory, methods, educational applications) around the latent semantic analysis
8	Article	K. VanLehn, P. Jordan, D. Litman (2007)	24	When Are Tutorial Dialogues More Effective Than Reading?	Compared whether natural-language-based computer tutoring was more effective than listening to a lecture or reading a text in seven experiments. The results depended on whether the students' preparation matched the content of the instruction

Table 5 (continued)

No	Type	Authors	Co-citation frequency	Title	Findings
9	Article	A. C. Graesser, D. S. Menamara, M. M. Louwerse, Z. Cai (2004)	23	Coh-Metrix: Analysis of text on cohesion and language	Introduced a computer tool called Coh-Metrix, which analyzed texts on over 200 measures of cohesion, language, and readability
10	Article	K. R. Koedinger, A. T. Corbett, Charles Perfetti (2012)	23	The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning	Proposed a KLI framework for analysis and evaluation by putting forward three coordinated classification methods, namely knowledge, learning, and teaching
11	Article	J. R. Anderson, A. T. Corbett, K. R. Koedinger, & Ray Pelletier (1995)	22	Cognitive tutors: Lessons learned	Presented empirical studies on testing the learning effects, elaborated how the procedure formed when the cognitive model was incorporated into a tutoring system, and how to build individual instruction
12	Article	V. Aleven, B. M. McLaren, J. Sewall, & K. R. Koedinger (2009)	22	A New Paradigm for Intelligent Tutoring Systems: Example-Tracing Tutors	Introduced the Cognitive Tutor Authoring Tools (CTAT) which could create example-tracing tutors without programming. The CTAT implements a simpler method that is more pragmatic for non-programmers
13	Article	A. C. Graesser, P. Chipman, B. C. Haynes, & A. Olney (2005)	20	AutoTutor: An intelligent tutoring system with mixed-initiative dialogue	Reported that AutoTutor obtained learning gains of approximately 0.8 sigma
14	Article	C. Conati, & H. Maclaren (2009)	20	Empirically building and evaluating a probabilistic model of user affect	Presented a probabilistic model of user affect relying on an emotion theory "OCC" model and designed to recognize users' emotions during human-computer interaction
15	Review	T. Murray (1999)	20	Authoring intelligent tutoring systems: An analysis of the state of the art	Summarized and categorized the research and development state of ITS systems

red circle, and the thickened lines in the larger red circle represent the citation relationships between the co-cited references with high Burst values from 2019 to 2020. The latest co-cited reference clusters with high Burst values were mainly in clusters #1 and #2, the themes of which reflect the emerging ITS research trends.

In the #2 cluster, the top terms were “problem-centered instruction” and “STEM,” for which there were three

The first identified paper was the “Effectiveness of Cognitive Tutor Algebra I at Scale” (Pane et al., 2014) (burst=4.15), which reported on a two year study of Cognitive Tutor Algebra I (CTAI) to compare the learning effects of personalized, mastery learning, blended learning, and CTAI alone. It was found that there was no noticeable learning effect in the first year, but positive effects were found in the second year. A second prominent paper was “Deep knowledge tracing” (Piech et al., 2015) (burst=4.05), which introduced Recurrent Neural Networks (RNNs) to assess large scale online teaching environments and student modeling. The RNN model family was found to more easily capture the complex representations of student knowledge than earlier models and was able to substantially improve student performance predictions. The third paper was “Stupid Tutoring Systems, Intelligent Humans” (Baker, 2016) (burst=4.02), which examined the importance of human intelligence in ITS research from a critical perspective, proposed a new teaching paradigm based on educational data mining and learning analytics methods, and emphasized the value of human wisdom when developing ITS applications.

Of the top five references with the strongest citation bursts in 2019–2020 in clusters #1 and #2, there were three meta-analytic reviews: “Intelligent Tutoring Systems and Learning Outcomes: A Meta-Analysis” by Ma et al. (2014) (burst = 10.16); the “Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review” by Kulik (2015) (burst = 9.1); and “A meta-analysis of the effectiveness of intelligent tutoring systems on college students’ academic learning. *Journal of Educational Psychology*” by Steenbergen-Hu and Cooper, (2014) (burst = 9.1). These references analyzed the effectiveness of the ITS-related learning effect evaluations in different environments.

Kulik (2015) conducted a meta-analysis of 50 controlled intelligent computer tutoring system evaluations and found that the intelligent tutoring had raised test scores by 0.66 standard deviations over the conventional level. Ma et al. (2014) examined 107 findings from 73 separate reports, finding that the average ITS effect resulted in an improvement in test scores of 0.43 standard deviations, but also found that the improvements depended to a great extent on whether they were measured using locally developed or standardized tests. The meta-analytic review by Steenbergen-Hu and Cooper (2014) analyzed 35 ITS effectiveness evaluations in colleges, finding that the ITS applications resulted in increases in the overall test scores by approximately 0.35 standard deviations, but that the

type of control group strongly influenced the evaluation results.

The other two references with the strongest citation bursts were a research article and a book. The article “Coh-Metrix: Providing Multilevel Analyses of Text Characteristics” (Graesser et al., 2011) (burst = 13.09) identified “word concreteness,” “syntactic simplicity,” “referential cohesion,” “causal cohesion,” and “narrativity” as the main factors accounting for the most variance in texts across grade levels and text categories using Coh-Metrix. The book “Automated Evaluation of Text and Discourse with Coh-Metrix” (McNamara et al., 2014) (burst = 12.37) provided a comprehensive introduction to Coh-Metrix from both theoretical and practical perspectives, and commented that the development of Coh-Metrix had resulted in a new paradigm that integrated language, corpus analyses, computational linguistics, education, and cognitive science research.

The top ten co-cited references with high burst values from 2019 and the first half of 2020 are listed in Table 6. As there was a large number of reviews (7 reviews, 2 articles, and 1 book), the proportion of original research has been relatively small. It showed that scholars in this field paid more attention to the review in this period. Generally speaking, reviews do not produce new knowledge as they only analyze previous studies, which seems to indicate that there have been fewer original research breakthroughs in

Table 6 Top 10 co-cited references with high burst value in 2019–2020.5

No	Type	Authors	Citation bursts	Title	Cluster ID
1	Review	K. VanLehn (2011)	16.83	The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems	#3
2	Article	A. C. Graesser, D. S. McNamara, J. M. Kulikowich (2011)	13.09	Coh-Metrix: Providing Multilevel Analyses of Text Characteristics	#1
3	Book	D. S. McNamara, A. C. Graesser, P. M. McCarthy, & Z. Cai (2014)	12.37	Automated Evaluation of Text and Discourse with Coh-Metrix	#1
4	Review	W. Ma, O. O. Adesope, J. C. Nesbit, & Q. Liu (2014)	10.16	Coh-Metrix: Capturing Linguistic Features of Cohesion	#2
5	Review	J. A. Kulik (2015)	9.1	Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review	#2
6	Review	S. Steenbergen-Hu, & H. Cooper (2014)	8.24	A meta-analysis of the effectiveness of intelligent tutoring systems on college students’ academic learning	#2
7	Article	K. R. Koedinger, A. T. Corbett, C. Perfetti (2012)	9.59	The Knowledge-Learning-Instruction Framework: Bridging the Science-Practice Chasm to Enhance Robust Student Learning	#14
8	Review	B. D. Nye, A. C. Graesser, & X. Hu (2014)	7.61	AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring	#15
9	Review	S. Steenbergen-Hu, & H. Cooper (2013)	5.88	A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students’ mathematical learning	#2
10	Review	M. C. Desmarais, & R. S. Baker (2012)	5.82	A review of recent advances in learner and skill modeling in intelligent learning environments	#2

recent years or any breakthroughs there have been have not received widespread attention. This situation was further confirmed when compared with Table 5, in which the top co-cited references in history were research articles. Liu and Hu (2018) believed that as there had been wide ranging discussions on education ITS developments, there was a great deal of high-level research repetition but few innovative breakthroughs. This was possibly because the ITS field has lacked a set of adaptive theoretical frameworks and basic guiding theories or assumptions, which should not be broad and descriptive, but directly reflect the characteristics and the essences of ITS-associated learning and education (Liu & Hu, 2018). Thus, we could boldly infer that there was little obvious breakthrough in ITSs in recent years, and there might exist a high level of research repetition.

Conclusion

Publication growth patterns and top contributors

In response to RQ 1, it was revealed that there had been a fluctuating growth in yearly ITS publications, and that the number of citations had exponential growth, which indicated that the ITS field was still in its initial developmental stage based on the Price literature exponential growth curve. However, high developmental potential was evident, which indicated that the ITS and ITS research is expected to flourish in the coming years.

It was implied that the milestones in the development of AI led to the three-stage development of ITSs research. In 1963, ITSs were first proposed and after the Lighthill report (1973) was published, research into AI was reduced with the study into ITSs commensurately reducing until 1985, as shown in Fig. 1. The development of the BP algorithm made the training of large scale neural networks possible, which resulted in increased interest in AI research and a renewed interest in ITSs. However, AI research experienced a second tough time in the late 1980s and early 1990s, then the research recovered. When Deep Blue defeated the world chess champion in 1997, ITS research increased. When Hinton's Deep Belief Net made a breakthrough in AI algorithms in 2006, there was an explosion in ITS research publications.

The most productive country/region for ITS research has been the USA, the most productive author has been Graesser, the top publication journal has been *Computers & Education* in terms of the number of the publications and the University of Memphis has contributed the most ITS research studies.

The multidisciplinary integration of ITSs

A new viewpoint was introduced to explain the subdiscipline composition to answer RQ 2. It was found that computer science, education research, psychology, and engineering have been the main ITS research knowledge sources, with computer science taking the dominant position. Due to its application-oriented characteristics, ITSs have had a unique discipline integration feature, which means that ITS research has tended to harness the latest achievements from other disciplines (Shneider, 2009). Given the continuing rapid technological developments, several ITS research subdivisions emerged associated with educational theory, cognitive psychology, and linguistics. This study explicated the path of multidisciplinary integration of ITSs researches in social and natural science fields. Social science in ITSs developed along with natural science and technology evolution.

The category analyses also clearly revealed ITS's multidisciplinary integration path. While ITSs originated from an educational discipline, since 1986, computer science research has become its most important knowledge base; however, from 2007, social science-based publications have exceeded natural science-based publications.

The most popular research issues

The top five keywords from the co-occurrence frequencies were identified as interactive learning environments, student modeling, teaching/learning strategies, machine learning, human-computer interface, and Coh-Metrix), which reflected the most popular issues in ITS research.

Intellectual base and emerging research trends of ITSs

The intellectual base evolution was analyzed by examining the co-cited references, the clusters for which were divided into three chronological research stages: the first stage mainly focused on the ITS applications; the second stage led to various subfields; human-computer interaction, probabilistic models, computer-mediated communication, cognitive tutors/meta-cognitive skills, semantic web-based educational systems/latent semantic analysis, STEM learning, data mining, and computer-supported collaborative learning; and the third stage gave rise to two main research subfields; computational linguistics, and the combination of computer technology and pedagogy.

The latest Co-cited references clusters with high Burst value references were clusters #1 and #2, the themes of which reflected the emerging ITS research trends. Cluster #1 was mainly focused on Coh-Metrix, indicating that the text analysis/NLP and other research around Coh-Metrix was the current research trend. Cluster #2 was focused on

problem-centered instruction and STEM learning, which have been traditional ITS research concerns. Model construction and critical thinking for the development of ITS applications were also found to be important recent research themes.

From the scientometrics view, it was argued that ITS research is currently multidisciplinary and therefore researchers, should implement ideas and practice high-risk endurance when choosing research tasks (Shneider, 2009).

Implication

Here we wish to raise a question for future research. What has been the effect of ITS on education? To answer this question, we need to consider the effects and functions that ITSs are seeking to achieve, the extent to which technology can be applied, as well as identifying the initial desire for ITSs from an educational viewpoint. This question comes from the vision of education. It doesn't need to be answered immediately, and there might be more questions to be addressed, especially under the situation of the global epidemic COVID-19 bringing great challenges to education. There are still many students taking classes at home around the world. What are the new requirements in the circumstances for ITSs? We hope to raise these questions to inspire researchers, educators to explore more undeveloped research areas, and have more in-depth discussions.

Limitations and future research

This article had some limitations as the data were only collected from the SCIE/SSCI indices in the WoS database. Although the WoS database contains a large volume of high impact research articles, some valuable research has been also published in books or collected in other indices. Diversified data sources should therefore be considered in the future. This study also excluded reviews from the data collection. In the future study, the trends of review research including systematic review, literature review and meta-analysis could be conducted.

References

- Ahuja, N. J., & Sille, R. (2014). A critical review of development of intelligent tutoring systems: Retrospect, present and prospect. *International Journal of Computer Science Issues (IJCSI)*, 10(4), 39.
- Akbulut, Y., & Cardak, C. S. (2012). Adaptive educational hypermedia accommodating learning styles: A content analysis of publications from 2000 to 2011. *Computers and Education*, 58(2), 835–842.
- Aleven, V., McLaren, B. M., Roll, I., & Koedinger, K. R. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16(2), 101–128.
- Aleven, V., McLaren, B. M., Sewall, J., & Koedinger, K. R. (2009). A new paradigm for intelligent tutoring systems: Example-tracing tutors. *International Journal of Artificial Intelligence in Education*, 19(2), 105–154.
- Almasri, A., Ahmed, A., Al-Masri, N., Abu Sultan, Y., Mahmoud, A. Y., Zaqout, I., Akkila, A. N., & Abu-Naser, S. S. (2019). Intelligent tutoring systems survey for the period 2000–2018. *International Journal of Academic Engineering Research (IJAER)*, 3(5), 21–37.
- Anderson, J. R. (1980). *Cognitive psychology and its implications*. San Francisco: Freeman.
- Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, Mass: Harvard University Press.
- Anderson, J. R., Boyle, C. F., Corbett, A. T., & Lewis, M. W. (1990). Cognitive modeling and intelligent tutoring. *Artificial Intelligence*, 42(1), 7–49.
- Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, 4(2), 167–207.
- Anwar, M. A., Zhou, R., Asmi, F., Wang, D., & Hammad, A. (2019). Mapping the evolution of energy-growth nexus: Synergies and trade-offs. *Journal of Economic Surveys*, 33(3), 1–31.
- Arroyo, I., Cooper, D. G., Burleson, W., Woolf, B. P., Muldner, K., & Christopherson, R. (2009). *Emotion sensors go to school*. IOS Press.
- Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition—implications for the design of computer-based scaffolds. *Instructional Science*, 33(5–6), 367–379.
- Baker, R. S., D'Mello, S. K., Rodrigo, M. M. T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive-affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies*, 68(4), 223–241.
- Baker, R. S., & Ryan, S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, 26(2), 1–15.
- Baker, R. S. J. D., Corbett, A. T. & Aleven, V. (2008). More accurate student modeling through contextual estimation of slip and guess probabilities in bayesian knowledge tracing. International Conference on Intelligent Tutoring Systems. Springer-Verlag.
- Beal, C. R., Arroyo, I. M., Cohen, P. R., & Woolf, B. P. (2010). Evaluation of animal watch: An intelligent tutoring system for arithmetic and fractions. *Journal of Interactive Online Learning*, 9(1), 64–67.
- Braam, R. R., Moed, H. F., & Van Raan, A. F. (1991). Mapping of science by combined co-citation and word analysis. Structural aspects. *Journal of the American Society for Information Science*, 42(4), 233–266.
- Cavanagh, P. (1963). THE AUTOTUTOR AND CLASSROOM INSTRUCTION 3 COMPARATIVE STUDIES INTRODUCTION. *Occupational Psychology*, 37, 44–48.
- Cavanagh, P. (1964). THE AUTOTUTOR AND CLASSROOM INSTRUCTION 3 COMPARATIVE STUDIES2 THE ROYAL AIR-FORCE STUDY. *Programmed Learning*, 1(1), 26–31.
- Chang, M., D'Aniello, G., Gaeta, M., et al. (2020). Building ontology-driven tutoring models for intelligent tutoring systems using data mining. *IEEE Access*, 8(1), 48151–48162.
- Chang, M., D'Aniello, G., Gaeta, M., Orciuoli, F., Sampson, D., & Simonelli, C. (2020). Building ontology-driven tutoring models for intelligent tutoring systems using data mining. *IEEE Access*, 8, 48151–48162.

- Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: A narrative overview. *Procedia Computer Science*, 136, 16–24
- Chen, C. (2005). The centrality of pivotal points in the evolution of scientific networks. *Proceedings of the 10th International Conference on Intelligent User Interfaces*, 98–105
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377
- Chen, C. (2012). Retrieved 29 Oct 2019 from <http://blog.sciencenet.cn/u/ChaomeiChen>
- Chen, C. (2017a). Science mapping: A systematic review of the literature. *Journal of Data and Information Science*, 2(2), 1–39
- Chen, C., Ibekwe-SanJuan, F., & Hou, J. (2010). The structure and dynamics of cocitation clusters: A multiple-perspective cocitation analysis. *Journal of the American Society for information Science and Technology*, 61(7), 1386–1409
- Chen, C., Song, I. Y., Yuan, X., & Zhang, J. (2008). The thematic and citation landscape of data and knowledge engineering (1985–2007). *Data and Knowledge Engineering*, 67(2), 234–259
- Chen, K., & Guan, J. (2011). A bibliometric investigation of research performance in emerging nanobiopharmaceuticals. *Journal of Informetrics*, 5(2011), 233–247
- Chen, W., Chan, T. W., Wong, L. H., Looi, C. K., Liao, C. C. Y., Cheng, H. N. H., & Pi, Z. (2020). IDC theory: Habit and the habit loop. *Research and Practice in Technology Enhanced Learning*. <https://doi.org/10.1186/s41039-020-00127-7>
- Chen, X., Li, J., Sun, X., & Wu, D. (2019). Early identification of intellectual structure based on co-word analysis from research grants. *Scientometrics*, 121(1), 349–369
- Chen, X., Xie, H., Zou, D., & Hwang, G.-J. (2020). Application and theory gaps during the rise of Artificial Intelligence in education. *Computers and Education: Artificial Intelligence*, 1, 100002. <https://doi.org/10.1016/j.caeai.2020.100002>
- Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science*, 25(4), 471–533
- Choi, B. C. K., & Pak, A. W. P. (2006). Multidisciplinarity, interdisciplinarity and transdisciplinarity in health research, services education and policy: 1 Definitions, objectives, and evidence of effectiveness. *Clinical and Investigative Medicine*, 29, 351–364
- Corbett, A. T., & Anderson, J. R. (1994). Knowledge tracing: Modeling the acquisition of procedural knowledge. *User Modeling and User-Adapted Interaction*, 4(4), 253–278
- Conati, C. (2009). Intelligent tutoring systems: New challenges and directions. *Proceedings of the Twenty-First International Joint Conference on Artificial Intelligence (IJCAI-09)*.
- Conati, C., & Kardan, S. (2013). Student modeling: supporting personalized instruction, from problem solving to exploratory open-ended activities. *Ai Magazine*, 34(3), 13–26.
- Conati, C., Gertner, A., & Vanlehn, K. (2002b). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12(4), 371–417
- Conati, C., Gertner, A., & Vanlehn, K. (2002a). Using Bayesian networks to manage uncertainty in student modeling. *User Modeling and User-Adapted Interaction*, 12, 371–417
- Conati, C., & Maclaren, H. (2009). Empirically building and evaluating a probabilistic model of user affect. *User Modeling and User-Adapted Interaction*, 19(3), 267–303
- Cotton, K. (2001). Classroom questioning. School improvement research series (SIRS). Retrieved 25 Oct 2005 from <http://www.nwrel.org/scpd/sirs/3/cu5.html>
- Craig, S., Graesser, A., Sullins, J., & Gholson, B. (2004a). Affect and learning: An exploratory look into the role of affect in learning with AutoTutor. *Journal of Educational Media*, 29(3), 241–250. <https://doi.org/10.1080/1358165042000283101>
- Dargue, B., & Biddle, E. (2014). Just enough fidelity in student and expert modeling for ITS. *International Conference on Augmented Cognition*. Springer International Publishing.
- Desmarais, M., & Baker, R. (2012a). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1–2), 9–38
- Desmarais, M., & Naceur, R. (2013). A Matrix factorization method for mapping items to skills and for enhancing expert-based Q-matrices. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial intelligence in education. AIED 2013. Lecture notes in computer science*. Springer.
- Desmarais, M. C., & d Baker, R. S. (2012). A review of recent advances in learner and skill modeling in intelligent learning environments. *User Modeling and User-Adapted Interaction*, 22(1), 9–38
- Deyi, L. (2018). *Introduction to artificial intelligence*. China Science and Technology Press.
- D’Mello, S., & Graesser, A. (2011). The half-life of cognitive-affective states during complex. *Cognition and Emotion*. <https://doi.org/10.1080/02699931.2011.613668>
- D’Mello, S., Graesser, A., & Picard, R. W. (2007). Toward an affect-sensitive autotutor. *IEEE Intelligent Systems*, 22(4), 53–61
- Du Boulay, B., Rebollo-Mendez, G., Luckin, R., Martínez-Mirón, E., & Harris, A. (2007). Motivationally intelligent systems: Diagnosis and feedback. In: AIED. 563–565.
- Egghe, L. (2005). Expansion of the field of informetrics: Origins and consequences. *Information Processing and Management*, 41(6), 1311–1316
- Elham, M., Nahid, Z., Sharareh, R. N. K., Mahnaz, R., Leila, K. & Marjan, G. S. (2018). Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2018.1558257>.
- Feng, F., Zhang, L., Du, Y., & Wang, W. (2015). Visualization and quantitative study in bibliographic databases: A case in the field of university–industry cooperation. *Journal of Informetrics*, 9(1), 118–134
- Gardner, H. (1987). The mind’s new science: A history of the cognitive revolution.
- Graesser, A., Chipman, P., & Leeming, F. (2009). Deep learning and emotion in serious games. *Serious games*. (pp. 105–124). Routledge.
- Graesser, A. C., Chipman, P., Haynes, B. C., & Olney, A. (2005a). Autotutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Transactions on Education*, 48(4), 612–618
- Graesser, A. C., Conley, M. W., & Olney, A. M. (2011). Intelligent tutoring systems. In S. Graham & K. Harris (Eds.), *APA educational psychology handbook: Vol 3. Applications to learning and teaching*. American Psychological Association.
- Graesser, A. C., & D’Mello, S. (2012). Emotions during the learning of difficult material. *Psychology of Learning and Motivation*, 57, 183–225
- Graesser, A. C., McNamara, D. S., Cai, Z., Conley, M., Li, H., & Pennebaker, J. (2014). Coh-Metrix measures text characteristics at multiple levels of language and discourse. *Elementary School Journal*, 115, 211–229
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-metrix: Providing multilevel analyses of text characteristics. *Educational Researcher*, 40(5), 223–234
- Graesser, A. C., McNamara, D. S., Louwerse, M. M., & Cai, Z. (2004). Coh-metrix: Analysis of text on cohesion and language. *Behavior Research Methods Instruments and Computers*, 36(2), 193
- Graesser, A.C., Moreno, K., Marineau, J., Adcock, A., Olney, A., Person, N., & The Tutoring Research Group (2003). AutoTutor

- improves deep learning of computer literacy: Is it the dialogue or the talking head? In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of artificial intelligence in education* (pp. 47–54). Amsterdam: IOS Press.
- Graham, S., Hebert, M., & Harris, K. R. (2015). Assessment and writing formative: A meta-analysis. *The Elementary School Journal*, 4(115), 523–547.
- Griffith, B. C., Small, H. G., Stonehill, J. A., & Dey, S. (1974). The structure of scientific literatures II: Toward a macro-and micro-structure for science. *Science Studies*, 4(4), 339–365.
- Hinton, G. E., Osindero, S., & Teh, Y. (2006). A fast learning algorithm for deep belief nets. *Neural Computation*, 18, 1527–1554.
- Hood, W. W., & Wilson, C. S. (2001). The literature of bibliometrics, scientometrics, and informetrics. *Scientometrics*, 52, 291.
- Hou, J., Yang, X., & Chen, C. (2018). Emerging trends and new developments in information science: A document co-citation analysis (2009–2016). *Scientometrics*, 115, 869–892.
- Hu, Z., & Zhou, T. (2017). New media industries from frontiers artificial intelligence and virtual reality. *Social Sciences Academic Press (China)*, 1, 2–3.
- Hu, X., Liu, F., & Bu, C. (2020). Research advances on knowledge tracing models in educational big data. *Journal of Computer Research and Development*, 57(12), 2523–2546.
- Isotani, S., & Mizoguchi, R. (2008). Theory-driven group formation through ontologies. *Intelligent Tutoring Systems*, 9th International Conference, ITS 2008, Montreal, Canada, June 23–27, 2008, Proceedings. DBLP.
- Johnson, W. L., & Rickel, J. W. (2000). Animated pedagogical agents face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education*, 11, 47–78.
- Khasseh, A. A., Soheili, F., Moghaddam, H. S., & Chelak, A. M. (2017). Intellectual structure of knowledge in iMetrics: A co-word analysis. *Information Processing and Management*, 53(3), 705–720.
- Koedinger, K. R., & Corbett, A. T. (2006). Cognitive tutors: Technology bringing learning science to the classroom. In K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences*. (pp. 61–78). Cambridge University Press.
- Koedinger, K. R., Corbett, A. T., & Perfetti, C. (2012). The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. *Cognitive Science*, 36(5), 757–798.
- Kolodner, J. (2002). Facilitating the learning of design practices: Lessons learned from an inquiry into science education. *Journal of Industrial Teacher Education*, 39(3), 9–40.
- Kort, B., Reilly, R., & Picard, R. W. (2002). An affective model of interplay between emotions and learning: reengineering educational pedagogy-building a learning companion. *Proceedings IEEE International Conference on Advanced Learning Technologies*. IEEE.
- Kulik, J. A. (2015). Effectiveness of intelligent tutoring systems: A meta-analytic review. *Review of Educational Research*, 86(1), 42–78.
- Kurshan, B. (2016). *The future of artificial intelligence in education*. Forbes Magazine.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Lawrence Erlbaum Associates.
- Lester, J. C., Ha, E. Y., Lee, S. Y., Mott, B. W., Rowe, J. P., & Sabourin, J. (2013). Serious games get smart: Intelligent game-based learning environments. *AI Magazine*, 34(4), 31–45.
- Leydesdorff, L., & Rafols, I. (2012). Interactive overlays: A new method for generating global journal maps from Web-of-Science data. *Journal of Informetrics*, 6(2), 318–332.
- Li, J., & Chen, C. (2016). *CiteSpace: Text mining and visualization in scientific literature*. Capital University of Economics and Business Press.
- Lighthill, J. (1973). Artificial Intelligence: A general survey. *Artificial Intelligence: a paper symposium*, Science Research Council.
- Liu, K., & Jing, Hu. (2018). The theory framework of AIED: The symmetric hypothesis between learner and educational resources—An interview with ITS expert professor Xiangen Hu[J]. *Open Education Research*, 24(06), 4–11.
- Liu, Q., Chen, E. H., Huang, Z. Y., Wu, R. Z., Su, Y., & Hu, G. P. (2008). Research on educational data mining technology for online intelligent learning. *Pattern Recognition and Artificial Intelligence*, 31(01), 77–90.
- Liu, Y. (2003). Modern educational technology and intelligent computer-aided teaching. *Journal of Natural Sciences of Harbin Normal University*, 2003(5), 59–61.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence Unleashed. An argument for AI in education*. Pearson.
- Ma, W., Adesope, O. O., Nesbit, J. C., & Liu, Q. (2014). Intelligent tutoring systems and learning outcomes: A meta-analysis. *Journal of Educational Psychology*, 106(4), 901–918.
- Madani, F., & Weber, C. (2016). The evolution of patent mining: Applying bibliometrics analysis and keyword network analysis. *World Patent Information*, 46, 32–48.
- Matz, M. (1981). *Towards a generative theory of high school algebra errors. Intelligent tutoring systems: An overview*. (p. 7). Academic Press.
- McCoy, J., Treanor, M., Samuel, B., Wardrip-Fruin, N., & Mateas, M. (2011). Comme Il Faut: A system for authoring playable social models. In *Proceedings of the Seventh International Conference on Artificial Intelligence and Interactive Digital Entertainment*, 158–163. Palo Alto, CA: AAAI Press.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- McNamara, D. S., Louwerse, M. M., McCarthy, P. M., & Graesser, A. C. (2010). Coh-Metrix: Capturing linguistic features of cohesion. *Discourse Processes*, 47(4), 292–330.
- Miranda, R., & Garcia-Carpintero, E. (2018). Overcitation and overrepresentation of review papers in the most cited papers. *Journal of Informetrics*, 12(4), 1015–1030.
- Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education*, 10, 98–129.
- Nalimov, V. V., & Mulchenko, Z. M. (1971). Measurement of science. Study of the development of science as an information process.
- Núñez, R., Allen, M., Gao, R., Rigoli, C. M., Relaford-Doyle, J., & Semenuks, A. (2019). What happened to cognitive science? *Nature Human Behaviour*, 3, 782–791.
- Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and family: A review of 17 years of natural language tutoring. *International Journal of Artificial Intelligence in Education*, 24, 427–469.
- Pane, J. F., Griffin, B. A., McCaffrey, D. F., & Karam, R. (2014). Effectiveness of cognitive tutor algebra I at scale. *Educational Evaluation and Policy Analysis*, 36(2), 127–144.
- Person, N. K., Graesser, A. C., Bautista, L., Mathews, E. C., & The Tutoring Research Group. (2001). Evaluating student learning gains in two versions of AutoTutor. In J. D. Moore, C. L. Redfield, & W. L. Johnson (Eds.), *Artificial intelligence in education: AI-ED in the wired and wireless future*. (pp. 286–293). Amsterdam: IOS Press.
- Piech, C., Spencer, J., Huang, J., Ganguli, S., Sahami, M., Guibas, L., et al. (2015). Deep knowledge tracing. *Computer Science*, 3(3), 19–23.

- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533–536
- Santhanam, R., Liu, D., & Shen, W. C. (2016). Research note gamification of technology-mediated training: Not all competitions are the same. *Information Systems Research*, 27(2), 453–465
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85–117.
- Sharma, S., Ghorpade, S., Sahni, A., & Saluja, N. (2014). Survey of intelligent tutoring systems: A review on the development of expert/intelligent tutoring systems, various teaching strategies and expert tutoring system design suggestions. *International Journal of Engineering Research and Technology*, 3(11), 37–42
- Shneider, A. M. (2009). Four stages of a scientific discipline; four types of scientist. *Trends in Biochemical Sciences*, 34(5), 217–223
- Shute, V. J., & Psotka, J. (1996). Intelligent tutoring systems: Past, present, and future. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology*. (pp. 570–600). Macmillan.
- Sleeman, D. H., & Brown, J. S. (1982). *Intelligent tutoring systems: An overview*. (pp. 1–11). New York: Academic Press.
- Small, H., & Griffith, B. C. (1974). The structure of scientific literatures I: Identifying and graphing specialties. *Science Studies*, 4, 17–40
- Spector, J., Merrill, M., & David, M. (2014). *Handbook of research on educational communications and technology*. Springer.
- Steenbergen-Hu, S., & Cooper, H. (2014). A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. *Journal of Educational Psychology*, 106(2), 331–347
- Steenbergen-Hu, S., & Cooper, H. (2013). A meta-analysis of the effectiveness of intelligent tutoring systems on K–12 students' mathematical learning. *Journal of Educational Psychology*, 105(4), 970
- Tague, J., Beheshti, J., Rees-Potter L. (1981). The law of exponential growth: Evidence, implications and forecasts. *LIBRARY TRENDS*.125–149.
- Tsay, C. H., Kofinas, A. K., & Luo, J. (2018). Enhancing student learning experience with technology-mediated gamification: An empirical study. *Computers in Education*, 121, 1–17
- Van Eck, N., & Waltman, L. (2009). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523–538
- Vandewaetere, M., Desmet, P., & Clarebout, G. (2011). The contribution of learner characteristics in the development of computer-based adaptive learning environments. *Computers in Human Behavior*, 27, 118–130
- VanLehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16, 227–265
- VanLehn, K. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197–221
- VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, 31, 3–62
- Wang, H. (1997). SQL Tutor+: A co-operative ITS with repository support. *Information and Software Technology*, 5(39), 343–350
- Wei, R. (2011). VOSviewer. Retrieved 8 Oct 2019 from <http://blog.sciencenet.cn/blog-113146-451966.html>
- Wenger, E. (1987). Artificial intelligence and tutoring systems. San Francisco CA Morgan Kaufmann Publish.
- Wescourt, K. T., Beard, M., GOuld, L., & Barr, A. (1977). Knowledge-Based CAI: CINs for Individualized Curriculum Sequencing (No. TR-290). STANFORD UNIV CALIF INST FOR MATHEMATICAL STUDIES IN THE SOCIAL SCIENCES.
- Woolf, B. P. (2009). *Building intelligent interactive tutors*. Morgan Kaufman.
- Woolf, B. P., Chad Lane, H., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. Special issue on intelligent learning technologies. *AI Magazine*, 10, 66–84
- Woolf, B. P., Lane, H. C., Chaudhri, V. K., & Kolodner, J. L. (2013). AI grand challenges for education. *AI magazine*, 34(4), 66–84
- Wu, D., Xie, Y., Dai, Q., & Li, J. (2016). A systematic overview of operations research/management science research in Mainland China: Bibliometric analysis of the period 2001–2013. *Asia-Pacific Journal of Operational Research*, 33(06), 1650044
- Xie, P. (2015). Study of international anticancer research trends via co-word and document co-citation visualization analysis. *Scientometrics*, 105, 611–622
- Xu, G., Zeng, W., & Huang, C. (2009). Research on intelligent teaching system. *Research on Computer Applications*, 2009(11), 4019–4022
- Yu, H., & Riedl, M. O. (2012). A Sequential recommendation approach for interactive personalized story generation. In Proceedings of the Eleventh International Conference on Autonomous Agents and Multiagent Systems, (pp. 71–78). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.