Defining the boundaries between Artificial Intelligence in Education, Computer-Supported Collaborative Learning, Educational Data Mining, and Learning Analytics: A Need for Coherence

How to cite:

© 2020 Rienties, Køhler Simonsen and Herodotou

https://creativecommons.org/licenses/by/4.0/

Link(s) to article on publisher’s website:
http://dx.doi.org/doi:10.3389/feduc.2020.00128
Defining the Boundaries Between Artificial Intelligence in Education, Computer-Supported Collaborative Learning, Educational Data Mining, and Learning Analytics: A Need for Coherence

Bart Rienties1*, Henrik Køhler Simonsen2 and Christothea Herodotou1

1 Institute of Educational Technology, The Open University, Milton Keynes, United Kingdom, 2 Smartlearning, Copenhagen, Denmark

This review aims to provide a concise overview of four distinct research fields: Artificial Intelligence and EDucation (AIED), Computer-Supported Collaborative Learning (CSCL), Educational Data Mining (EDM), and Learning Analytics (LA). While all four fields are focused on understanding learning and teaching using technology, each field has a relatively unique or common perspective on which theoretical frameworks, methods, and ontologies might be appropriate. In this review we argue that researchers should be encouraged to cross the boundaries of their respective field and work together to address the complex challenges in education.

Keywords: artificial intelligence in education, computer-supported collaborative learning, educational data mining, learning analytics, review

INTRODUCTION

In the last 20 years a range of disciplines have been developed in the broad field of education and technology. Since the early 1980s the broad field of Artificial Intelligence and EDucation (AIED) emerged that aimed to use a combination of Artificial Intelligence (AI), learning theory, and educational practice to improve learning outcomes for learners using computers (Boyd et al., 1982; Holmes et al., 2019). Within AIED various subfields of research emerged based upon the power of computing and machine learning, such as intelligent tutoring systems (Aleven and Koedinger, 2002), adaptive hypertext systems (Eysink et al., 2009; Romero et al., 2009), and Computer-Supported Collaborative Learning (CSCL). Since the early 1990s a range of CSCL publications appeared exploring how learners and teachers could work together online using computers. A vast number of CSCL studies (e.g., Gunawardena, 1995; Roschelle and Koschmann, 1996; Fischer and Mandl, 2005; Rienties et al., 2009) have found that scaffolding, self-regulation, task design, and teaching presence are important concepts that can encourage learners to effectively work together.

In the mid-2000s a third stream of researchers (e.g., Baker and Yacef, 2009; Rosé et al., 2014) using Educational Data Mining (EDM) started to explore learning processes using bigger data sets and increased interconnections between data. Since 2011 a fourth research field of Learning Analytics (LA) emerged, which is specifically focused on understanding the complex
learning processes and learning outputs, using a multi-disciplinary combination of computer-science, educational psychology, engineering, and learning sciences (Ferguson, 2012; Papanisioti and Economides, 2014). In this contribution we aim to define what the potential boundaries and synergies are between AIED, CSCL, EDM, and LA, and how a combined interdisciplinary perspective can help to maximize the potential of these four research fields to understand the complexities of learning and teaching using technology. This might be particularly relevant for researchers and practitioners who may be new to these research fields. For a more detailed and deeper analysis of these fields, we encourage readers to connect to the respective journals in Table 1.

FOUR PERSPECTIVES ON COMPUTING, LEARNING, AND EDUCATION

The boundaries between AIED, CSCL, EDM, and LA are rather blurred. In part, this is because researchers and practitioners from these respective fields look at similar, yet slightly distinct phenomena, and in part, this is because researchers often work in interdisciplinary research groups across the boundaries of their specific research focus (Jeong et al., 2014; Aldowah et al., 2019; Dormezil et al., 2019). Therefore, the characterisations of the four research fields below are by definition an oversimplification of their complex, inter-linked, and fluid perspectives, relations, methodologies, and ontologies. Given that these fields emerged, faded, merged, and re-emerged at various points of time, rather than giving a historical overview of these fields, we will describe these fields in alphabetical order and in relation to the following aspects (see Table 1): (a) main aim/target, (b) educational and other underpinnings, (c) techniques and approaches, (d) society, and (e) conferences and journals.

Artificial Intelligence in Education

Although there is not a single definition of what AI might be, AI broadly refers to “computers which perform cognitive tasks, usually associated with human minds, particularly learning, and problem-solving” (Baker et al., 2019, p. 10). It is an umbrella term used to describe several methods such as machine learning, data mining (DM), neural networks or an algorithm (Zawacki-Richter et al., 2019). Its roots can be traced back to computer science and engineering, with a strong relation to economics, cognitive science, philosophy, and neuroscience (Popenici and Kerr, 2017; Holmes et al., 2019; Zawacki-Richter et al., 2019). As indicated in Table 1, the main aim of AIED is to simulate and predict learning processes. In terms of philosophical underpinning, a crucial underlying assumption of AI, and AIED in particular, is that any aspect of learning or any other feature of intelligence can be described, and that a machine is able to simulate it (Zawacki-Richter et al., 2019). In the last 20 years, substantial progress has been made in machine learning, which allows researchers to understand, model and simulate the complex behaviors of humans, which are assumed to be rational. Popenici and Kerr (2017, p. 2) defined machine learning “as a subfield of artificial intelligence that includes software able to recognize patterns, make predictions, and apply newly discovered patterns to situations that were not included or covered by their initial design.” With the incredible advances of AI in other sectors (e.g., automobile, health care, manufacturing), recently there has been a renewed interest in AIED (Tuomi, 2018; Zawacki-Richter et al., 2019).

For example, in a review of 146 studies conducted between 2007 and 2018 (Zawacki-Richter et al., 2019) a range of applications of AI in higher education were identified, including making admission decisions and course scheduling (Andris et al., 2013), assessment and feedback (Adamson et al., 2014), intelligence tutoring systems (Aleven and Koedinger, 2002), profile and prediction of students dropping out (Rizvi et al., 2019), and student models and academic achievement (Rizvi et al., 2019). As identified by Zawacki-Richter et al. (2019), although substantial progress has been made in AIED, most studies are quantitative in nature, make use of human intervention studies (Blanchard, 2012), with a control and experimental group, lack reflection on risks, challenges and ethical implications, and present a weak connection to relevant educational theories.

Computer-Supported Collaborative Learning

A main aim of CSCL is to understand the complex interactions in and outside class settings. While AIED assumes that all learning can be described and simulated by machines, in CSCL literature there is often a recognition that learning is complex, and socially constructed. McKeown et al. (2017, p. 439) argued that “(r)esearch in CSCL focuses on learning as a cognitive and/or social process and studies learning designs, learning processes, and pedagogic practices that support technology-mediated collaborative processes in communities of practice.” Given its focus on people working together, there are complex and dynamic interactions that may, or may not, be easily identifiable by computers (e.g., body language, cultural differences, emotions, linguistic styles). In order to develop and maintain a successful CSCL culture, Jeong et al. (2014) theorized that technology used for collaboration in CSCL needs to include: (1) a joint task, (2) communication, (3) sharing of resources, (4) engagement in productive processes, (5) engagement in co-construction, (6) monitoring and regulation, and (7) finding and building groups and communities. In face-to-face and blended learning scenarios, this maintenance of successful discourse might be difficult to achieve, while in online settings there is a wealth of research showing complexities in online collaboration (Fischer and Mandl, 2005; Renties et al., 2009). For example, in a review of 180 articles published in CSCL conferences in the period 2005–2017, Xia and Borge (2019) found that most studies focused on interaction in classrooms (47%), technology implemented in classrooms (13%), technology implemented in informal settings (15%), and in labs (11%). This strong focus on in-class analysis seems substantially different to AIED. Furthermore, CSCL seems to have strong experimental and learning science roots (Wis and Schwarz, 2017), whereby approximately half of recent studies identified by Jeong et al. (2014) used a methodologically strong...
design. At the same time, several meta reviews indicated a need for CSCL researchers to embrace more analytics and multi-level approaches to extend their methodological toolbox as well as the rigor of their studies beyond a single classroom or context (Jeong et al., 2014; Wise and Schwarz, 2017; Xia and Borge, 2019).

### Educational Data Mining

The main aim of EDM could be succinctly described as analyzing data from educational systems. With the rise of educational data, EDM has been going from strength to strength (Koedinger et al., 2015; Dutt et al., 2017; Aldawah et al., 2019). Early literature reviews (Romero and Ventura, 2007, 2010) noted the need for considering pedagogical aspects when mining data from educational systems, and identified benefits for students and teachers when recommender systems are used. Building on the first EDM conference in 2008, EDM has been defined (Baker and Yacef, 2009) as “an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in.” By using a range of DM techniques, EDM researchers aim to discover novel and potentially useful information from large amounts of data. As argued by a range of EDM researchers, while DM techniques are useful in big data contexts, in education there is a need to adjust algorithms to specific contexts (Dutt et al., 2017). Koedinger et al. (2015) explained that EDM focuses on a range of research questions in the psychology of learning: (a) assessment of cognition and learning, (b) transfer of learning, and discovery of cognitive models, (c) affect, motivation, and metacognition (Rosé et al., 2014), and (d) language and discourse analytics.

A desirable sequence of EDM research is to start off with DM leading to new statistical models of data, followed by building an (adaptive) automated system, and finally, closing the loop, by running an evidence-based experiment (Koedinger et al., 2015). In a review of 166 EDM studies, Dutt et al. (2017) identified five common clusters of studies: (1) analyzing student motivation, attitude and behavior; (2) understanding learning style; (3) e-learning; (4) collaborative learning; (5) EDM using clustering. A particular notable distinction between EDM, CSCL, and LA is the lack of specific reliance on educational theory. Most EDM research is considered pedagogically and educational theory-neutral, as the focus is on data discovery, testing of interventions, and optimizing models.

### Learning Analytics

The Journal of Learning Analytics defines LA as “... research into the challenges of collecting, analyzing, and reporting data with the specific intent to improve learning.” We define the main aim of LA as to improve learning processes. Several higher education institutions and distance learning providers have started to explore the use of LA dashboards that can display learner and learning behavior to teachers and instructional designers in order to provide more real-time or just-in-time support to students (Jivet et al., 2018; Herodotou et al., 2020). Furthermore, several institutions have developed predictive LA approaches to help identify, as early as possible, students who may be considered “at risk” of failing, and which of those students may need additional support (Viberg et al., 2018; Herodotou et al., 2020). Some institutions are also currently experimenting with providing LA data directly to students in order to support their learning processes and self-regulation (Winne, 2017; Rienties et al., 2019).

As argued by a range of authors, the distinction between EDM and LA is rather unclear, as leading researchers from both fields contribute to similar themes and debates across the two.
fields (Aldowah et al., 2019; Dormezil et al., 2019). According to Papamitsiou and Economides (2014), both EDM and LA communities share compatible goals and focus where learning science and data-driven analytics intersect. However, there are some subtle and more explicit differences in their ontological origins, techniques used, and perhaps most importantly the specific topics of interest. As argued by Papamitsiou and Economides (2014, p. 50) “LA adopts a holistic framework, seeking to understand systems in their full complexity. On the other hand, EDM adopts a reductionistic viewpoint by analyzing individual components, seeking for new patterns in data and modifying respective algorithms.” In a review contrasting 1,952 LA articles with 783 EDM articles by Dormezil et al. (2019), several common themes were identified, such as “educational computing” and “student performance.” LA focuses mostly on instruction and communication, student learning objectives and natural language processing. In contrast, EDM is focused on student performance and the technical specifications of respective predictive approaches, in particular "learning algorithms" and "student models." Nonetheless, there is more common overlap than distinct differences; Dormezil et al. (2019) argued that LA is probably best described as one domain with one prominent subset, that of EDM.

**DISCUSSION**

This review has briefly explored the intersection between education and technology in four fields: AIED, CSCL, EDM, and LA. In the last decade tremendous progress has been made to better understand the complexities of learning and teaching with technology. With the rise and availability of big data in education and AI, substantial leaps in the conceptual, theoretical, and evidence-based understanding of learning and teaching have been made in the four fields discussed. However, as highlighted by a range of reviews, most of these innovations have been localized in small lab studies, or in a single course, or specific context, with limited large-scale adoption within and across institutions (Viberg et al., 2018; Herodotou et al., 2020).

In order to truly make substantial leaps in the actual adoption of technology in large educational settings, achieve wide-spread uptake in educational institutions, and improve our understanding of the complexities of learning that can advance our theoretical models, we argue that the four research fields need to break down some of the artificial barriers between the respective communities, and jointly work together as one interdisciplinary research field. This can be achieved via a web of inter-related activities. First of all, national and international funding bodies should explicitly embrace and fund interdisciplinary research that cuts across the four (and other) fields. Second, by building cross-disciplinary network opportunities for researchers to learn from different disciplines might help to cross-fertilize and cross-pollinate different research ideas, methods and approaches. This can be “formally” achieved by including specific tracks in conference programs, joined special issues, and running some events together, as well as informally by encouraging research visits and invited seminars. Third, as highlighted in Table 1, there are substantial synergies that are possible in terms of theoretical, empirical and methodological advancement between the four fields. We argue that by bringing the best research minds together across the four fields, substantial progress can be made to address some of the large challenges in education and society at large. Toward this direction, in the last few years we have seen several initiatives that attempt to bring those fields closer, including the Festival of Learning and the creation of the International Alliance to Advance Learning in the Digital Era1 that brings the various societies included in Table 1 together. In terms of next steps following this work, and given the short-length nature of this article, a systematic and exhaustive review across the four fields would be particularly beneficial and help establish how exactly these fields differ and overlap.

**AUTHOR CONTRIBUTIONS**

All authors contributed to the article and approved the submitted version.

**FUNDING**

This article had received funding from the Horizon 2020 Research and Innovation Program ERASMUS+ (KA203-2019-002).

---

1http://www.alliancels.com/

---

**REFERENCES**


**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Rientes, Kahler Simonsen and Herodotou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.