Education Tech Research Dev
https://doi.org/10.1007/s11423-021-09977-4

DEVELOPMENT ARTICLE

Ethical issues in learning analytics: a review of the field

Dimitrios Tzimas[1] · Stavros Demetriadis[1]

Accepted: 22 February 2021

Abstract

Learning analytics (LA) collects, analyses, and reports big data about learners to optimise learning. LA ethics is an interdisciplinary field of study that addresses moral, legal, and social issues; therefore, institutions are responsible for implementing frameworks that integrate these topics. Many of the ethical issues raised apply equally to educational data sets of any size. However, in this study, we focus on big data that increases the scale and granularity of data gathered. The purpose of this study is twofold: (a) to critically review the published (2011–2018) scientific literature on LA ethics issues and (b) to identify current trends and answer research questions in the field. This study's research questions are as follows: what is essential in LA ethics for key educational stakeholders, and what should a proposed checklist for LA ethics include for specific educational stakeholders? After systematically searching online bibliographic databases, journals, and conferences, a literature review identified 53 articles from a sample of 562. The selected articles, based on critical and qualitative content analysis, were exhaustively analysed. The findings demonstrate the shortage of empirical evidence-based guidelines on LA ethics and highlight the need to establish codes of practices to monitor and evaluate LA ethics policies. Finally, this work proposes a useful checklist as an instructional design model for scholars, policymakers, and instructional designers, so that trusted partners may use LA responsibly to improve teaching and learning.

Keywords Learning analytics · Ethics · Systematic review · Checklist · Instructional design

Introduction

In this work, the researchers conducted a systematic literature review (Kitchenham, 2004) to map the critical concerns associated with learning analytics (LA) ethics. The authors expect this qualitative content analysis to be a synopsis and guide for key educational stakeholders who wish to gain insights into this emerging field. This work's main contribution is analysing how LA's ethical issues are defined, analysed, and resolved.

✉ Dimitrios Tzimas
detzimas@csd.auth.gr

Stavros Demetriadis
sdemetri@csd.auth.gr

[1] School of Informatics, Aristotle University, Thessaloniki, Greece

The following sections provide a description of LA ethics and an overview of relevant work. The "Background" section presents data analytics, LA, and LA ethics before establishing the research questions. Then, the subsequent three sections are concerned with the research design and the literature review's execution. The "Method" section presents the methodology, referencing the bibliographic databases and the selection criteria that resulted in the final 53 selected publications. The "Results" section presents insights into the literature review and proposes a mnemonic checklist with data management methods for best practice in establishing trustworthy LA ethics. The "Discussion" section discusses the aforementioned insights, comments on the next steps, and suggests future research's potential direction.

Background

Learning analytics

It is the epoch of big data, social networks, and cloud computing. Every piece of data is captured and leaves a digital trail (Siemens & Long, 2011), "increasing the volume, variety, velocity and veracity of student data" (Prinsloo & Slade, 2017, p. 8). Big data analytics thus involves using sophisticated analytical techniques of processing large datasets that produce useful conclusions to improve the organisational purposes and learner-centred, customised paradigm of education (Asamoah et al., 2017; Jantti & Heath, 2016). Big data refers to data gathering, data analytics, and decision making based on analytics.

The Society for Learning Analytics Research (SoLAR) (Lang et al., 2017) defines LA as "the measurement, collection, analysis, and reporting of big data about learners and their contexts and behaviours, for purposes of understanding and optimising learning and the environments in which it occurs" (Siemens & Long, 2011, p. 33). Learning analytics combines information retrieval, machine learning (ML), data visualisation, and statistical algorithms (Siemens, 2012). In addition, LA, as a big data practice, is an interdisciplinary field that uses concepts from computer science, statistics, behavioural science, instructional theory, and educational psychology to provide benefits for students, instructors, and institutions. Importantly, LA is related to learning sciences, as a learning theory can transform information from learning analytics into actionable knowledge for instructional design (Wong et al., 2019). Furthermore, LA is related to business, web, academic, action, and predictive analytics (Papamitsiou & Economides, 2014), and it produces information that instructors can translate into action (Prinsloo & Slade, 2017). Support for LA may include self-assessments, recommender systems, visualisations, personalised learning paths, and real-time feedback (Ifenthaler & Schumacher, 2016). Finally, student-facing LA supports institutions in resource allocation, student success, and finance (Leitner et al., 2017).

Learning analytics ethics

Ethics is a framework of moral principles concerned with what is right for individuals and society (Gray & Boling, 2016). In the literature, ethics, as a moral practice, is defined as the systematisation of correct and incorrect behaviour. Ethics is a branch of philosophy that began with the ancient Greek philosophers, mainly Aristotle and Socrates. A much later application of ethics is professional ethics. The main types of professional ethics are deontological, consequentialist, virtue, and applied ethics. Deontology, as a rule-based perspective, “addresses human action itself, assuming that an act itself can be seen to be inherently good, insofar as it references a formalized set of rules” (Gray & Boling, 2016, p. 4). In contrast to deontological ethics, consequentialism decides whether an act is right by considering its consequences. Virtue ethics is an approach that emphasises an individual’s character as a vital element in ethical thinking.

Cardinali et al. (2015) defined ethics as a moral code of norms that exist in society externally to a person, depending on culture and time, while ethical decision making is considerably challenging and complex (Spector, 2016). Moreover, Drachsler and Greller (2016) stated that the Nuremberg Code provided the first written ethical research principles. Ifenthaler and Tracey (2016) identified that legal considerations and ethical issues are related to the use of educational data for LA. Furthermore, Siemens (2013) stated that students value the support that they receive from LA actions; however, they fear the bias of data-based decisions. Slade and Prinsloo (2013) and Wilson et al. (2017) compared LA with other big data contexts: the physical sciences, human resource management, national security, business intelligence, biomedicine, and public health. In this regard, Timmis et al. (2016) concluded that many ethical issues are not specific to LA but to other domains.

It is worth mentioning that many of the ethical issues raised apply equally to educational data sets of any size. However, in this study, we focus on big data that increases the scale and granularity of data gathered in current times.

Therefore, in this paper, we follow the rule-based consequentialism using concepts of the deontological (i.e., principles of action and duty to society) and consequentialist perspective. According to this approach, a human relies on chosen rules that define certain consequences.

Contradictions in the literature

The literature addresses many technological, pedagogical, and policy contradictions that stakeholders must face. This study analyses these antagonisms to advance the philosophical perspective by emphasising the antithetical perspectives vital to various stakeholders.

All stakeholders have viewpoints and commitments (Papa & Armfield, 2018; Reidenberg & Schaub, 2018), and an imbalance of power is, therefore, possible between stakeholders (Jones & Salo, 2018; Slade & Prinsloo, 2013). In this first contradiction, on the one hand, instructors have an ethical responsibility to use LA results to support their learners (Arnold & Sclater, 2017; Prinsloo & Slade, 2017) without this use resulting in harmful interventionism and paternalism (Jones, 2017; Scholes, 2016). On the other hand, as data producers and consumers, learners demand instruction in their learning procedures to gain benefits that encourage their engagement (Herder & Kawase, 2012) while also having fears associated with privacy (Ifenthaler & Schumacher, 2016). Institutions must strike a balance between the LA perspective (i.e., asking for more student data) and the student perspective (i.e., asking for data limitation) (Pardo & Siemens, 2014). Furthermore, decision-makers use deterministic, data-driven algorithms that are based on behaviourism. In contrast, the learning phenomenon has a probabilistic dimension and requires stochastic models and specific learning theories (Fynn, 2016; Siemens, 2013).

Ethics differs around the world (Willis et al., 2016). Thus, a second contradiction refers to different viewpoints on ethical issues across countries; these varying viewpoints have created difficulty developing LA frameworks. For example, at the governance level, different laws between countries (Reidenberg & Schaub, 2018), the global market for LA tools, and various approaches between institutions (Cardinali et al., 2015) stakeholders should consider to achieve communication and efficiency. In other words, when moving solutions out of the labs, practitioners faced restrictions determined in national laws and justified in privacy frameworks (Hoel et al., 2017).

A third antagonism arises between the benefits and drawbacks of LA tasks. On the one hand, quality learning outcomes exist, such as self-regulation and personalised learning; however, on the other hand, issues such as surveillance and stereotypes are also present (Wintrup, 2017). Therefore, we must discuss learners' rights to know (Arnold & Sclater, 2017), be forgotten (Hoel et al., 2017), restrict processing, and opt-out (Sclater, 2016), and, in parallel, institutions' obligations to act and strive to support learners.

The final contradiction lies in technology and regulations. Technology and practice are developing quickly, producing tools and techniques that are rapidly changing trends, while legal frameworks are changing slowly; the legal system is immature concerning privacy and ethics concerns in analytics (Siemens, 2013). A further criticism in this area is that LA focuses on already existing data, while education and learning should enhance innovative ideas (Greller & Drachsler, 2012).

Learning analytics and instructional theory

We use as a conceptual organising framework instructional design theory (Reigeluth & Carr-Chellman, 2009). Design theories "are prescriptive in nature, in the sense that they offer guidelines as to what method(s) to use to best attain a given goal" (Reigeluth, 1999, p. 7). An instructional design hierarchy comprises the instructional "situation" at the top of the hierarchy that has "conditions" and "values" filling out the analysis elements and methods filling out the solution elements (Fig. 1). "Needs" are subordinate to the situation and are elicited during the instructional theory framework's conditions and values. Conditions describe the matters which could be empirically confirmed. Values are the elements of instruction that are matters of opinion about essential elements (i.e., learning goals, priorities for successful instruction, and methods) (Lin & Spector, 2017; Reigeluth & Carr-Chellman, 2009; Spector, 2015).

Our work accurately follows the architecture of instructional theory, specifically the practice of design layering (Gibbons & Rogers, 2009) that includes seven overlapping and interactive layers: content, strategy, message, control, representation, media logic, and data management. Our approach focuses on the strategy and data management layers. The strategy layer specifies the way to convey the content and stakeholders' roles and responsibilities. The instructional design data management layer specifies the data capture, storage,

Instructional situation (LA ethics)

1. Analysis elements
    - 1.1. Values (matters of opinion)
        - 1.1.1. Values about data management goals
        - 1.1.2. Values about power (stakeholders)
    - 1.2 Conditions (matters of fact)
        - 1.2.1 Instructional development constraints (legal/regulatory constructs)
2. Solutions elements
    - 1.3 Data management methods

Fig. 1 Constructs about the instructional theory framework that is aligned to the data management layer analysis, and interpretation that stakeholders can use to make instructional decisions (Gibbons, 2014).

In the LA cycle, when the instructional designer (i.e., the instructor or technological agent) decides to apply an intervention or revision to the learners' population, the intervention extracts (big) learning data. The collection, processing, analysis, presentation, and interpretation of these data in the data management layer must be made ethically (i.e., with LA ethics). Therefore, we should not apply an instructional design without using ethics as a filter in the data management layer. Consequently, we use instructional theory as the organising framework for selecting data management methods, which are methods for collecting and analysing learners' data ethically. Before defining the methods/prescriptions that represent the data management layer, we must analyse the situation. The situation which focuses on instructional data could suggest values about goals as statements about which ethical data practices are valued philosophically (Reigeluth & Carr-Chellman, 2009). Thus, ethics is aligned with the values-side of instructional theory ("matters of opinion"), precisely "values about goals." Legal and regulatory constructs could reflect conditions ("matters of fact"), specifically instruction development constraints. Consequently, LA ethics is expected to operate as a filtering device in the feedback loop provided by data management and analysis to the instructional designer. All relevant interventions to improve or adjust the design are compliant with the restrictions imposed by LA ethics. Overall, this article relates an instructional design for LA.

Figure 1 is a hierarchical diagram that summarises the constructs in describing the proposed instructional theory. It is similar to Fig. 1(1.2) on p. 24 of Reigeluth and Carr-Chellman (2009), except that it is aligned to the data management layer and selecting data methods.

Research questions

This work exhaustively studies, filters, compares the selected articles, and then extracts research questions and results. The following meaningful research questions (RQs) guide this review:

First main RQ: What is essential in learning analytics ethics for key educational stakeholders?

Second main RQ: What should a proposed checklist for learning analytics ethics include for specific educational stakeholders?

Method

Research design

For the systematic literature review undertaken in this study, extensive research of LA's literature was conducted from May 2017 to December 2018 to understand and document the current trends in the LA ethics subfield. Table 1 lists the literature that we searched and presents the journal titles and publications' distribution. We applied the following selection criteria: the search term "Learning Analytics" was used, and we performed the search in the abstracts, author keywords, and titles of the candidate articles. Searches were limited to articles published in English from 2011 (when LA was in the early stages of its implementation) to 2018 in journals and conference proceedings.

Table 1 Sources of the bibliographical research

Databases	IEEE Xplore Digital Library, Elsevier Digital Library through Scopus search engine, ScienceDirect, Wiley InterScience, Oxford University Press Digital Library, ACM digital library, and Springer
Journals and the number of studies (in descending order)	Educational Technology Research & Development (9), Educational Technology & Society (3), Journal of Learning Analytics (2), Computers in Human Behavior (1), Online Learning, IEEE Transactions on Learning Technologies, British Journal of Educational Technology, Journal of e-Learning and Knowledge Society, British Educational Research Journal, Higher Education Policy, The Information Society, International Journal of Technology Enhanced Learning, American Behavioral Scientist, Computers & Education, Technology, Knowledge and Learning. Theory and Research in Education, WIREs Data Mining and Knowledge Discovery, Research and Practice in Technology Enhanced Learning, International Review of Research in Open and Distributed Learning

With regard to the paper selection process, after systematically searching the above sources, 562 articles initially met the selection criteria. After studying their abstracts and conclusions, we finally selected a corpus of 53 papers that covered exclusively ethical principles for LA. The study of LA ethical issues was used as the criterion for selecting the final 53 papers. Only 9% of the articles considered ethics in relation to the conducted research. The articles that were studied comprised conceptual articles based on critical analysis, qualitative content analysis, and significant reviews of literature or empirical studies. The remaining articles related to LA in general but not to ethics were excluded from this work because they could not be used to answer this study's research questions. This research included only articles directly related to the research and presented factual information about LA ethics (Table 2). Afterwards, this study used spreadsheets without statistical methods to organise and analyse the data and findings, profiling the papers to create clusters with the LA ethics subfield's significant dimensions. In order to extract results, the articles were studied by considering the following four steps: analysis, selection, classification, and interpretation.

Our review's reliability and validity depend on conducting a consistent research design. Then, we used investigator triangulation to involve two authors' interpretation of the results.

Results

The ethical concerns with learning analytics (RQ1)

To answer the first research question (what is essential in LA ethics for key educational stakeholders?), we report the outcomes in the form of an unbiased list of instructional values about data management goals, including the significant LA ethical issues discussed in 53 reviewed articles. The analysis includes the following pattern strategy according to

Table 2 The articles that contributed to each issue and set of guidelines

Ethical issue (value about goals)	Articles	Guidelines emerging from the literature	References for guidelines
Privacy	Angeli et al. (2017); Avella et al. (2016); Beattie et al. (2008); Cruz et al. (2015); Drachsler and Greller (2016); Dyckhoff et al. (2012); Gray and Boling (2016); Greller and Drachsler (2012); Herder and Kawase (2012); Hoel et al. (2017); Ifenthaler and Schumacher (2016); Ifenthaler and Tracey (2016); Lawson et al. (2016); Manca et al. (2016); Scholes (2016); Siemens and Long (2011); West et al. (2016); Willis et al. (2016); Wintrup (2017), Reidenberg and Schaub (2018); Hoel and Chen (2018)	Co-operation between all stakeholders about privacy issues in data collection and analysis stages	Avella et al. (2016); Tsai and Gasevic (2017); Gursoy et al. (2017)
Openness and transparency	Willis et al. (2016); West et al. (2016); Timmis et al. (2016); Scholes (2016); Lawson et al. (2016); Ifenthaler and Tracey (2016); Ifenthaler and Schumacher (2016); Hoel et al. (2017); Herder and Kawase (2012); Greller and Drachsler (2012); Gray and Boling (2016); Dyckhoff et al. (2012); Cruz et al. (2015); Avella et al. (2016); Arnold and Sclater (2017); Taylor et al. (2018)	Clarity, data control, and accountability	Sclater (2016), Cardinali et al. (2015)
Labelling (autonomy)	Wintrup (2017); Siemens and Long (2011); Lawson et al. (2016); Gray and Boling (2016); Beattie et al. (2008); Wachter (2018)	Data will not be used to stereotype learners negatively	Beattie et al. (2008); Ari and Brandon (2014); Slade (2016); Lawson et al. (2016)
Resolve data ownership	West et al. (2016); Timmis et al. (2016); Lawson et al. (2016); Ifenthaler and Schumacher (2016); Hoel et al. (2017); Avella et al. (2016)	The schools' data must not be sold	Beattie et al. (2008); Cardinali et al. (2015); Lawson et al. (2016)

Table 2 (continued)

Ethical issue (value about goals)	Articles	Guidelines emerging from the literature	References for guidelines
Algorithmic fairness	Arnold and Sclater (2017); Beattie et al. (2008); Siemens (2013); Timmis et al. (2016); Fynn (2016)	Avoid biases to data interpretations	Hoel et al. (2017); Slade and Prinsloo (2013); Slade (2016); Willis et al. (2016)
Duty to act	Prinsloo and Slade (2017); Arnold and Sclater (2017)	Learners have the right to know	Slade (2016); Willis et al. (2013)

instructional design theory, as design involves the decomposition of a complete problem into subproblems of solvable size (Gibbons, 2014):

- Define/describe the instructional situation and values about goals to collect data ethically
- Determine the number of articles in which each value is discussed
- Identify major views/opinions

More specifically, our comparative analysis in a bottom-up approach to the selected literature resulted in findings illustrated in Fig. 2 for the situation that focuses on ethical data. This classification extracts a list of six key ethical dimensions representing values about goals associated with the data management layer. More specific, the suggested values are statements about which ethical data practices are valued philosophically. Thus, these coding categories (values about goals) describe the domain of ethics in LA to gain insights into the taxonomy of data management methods that enable LA to be ethical: Privacy; Transparency; Labelling; Data ownership; Algorithmic fairness; The obligation to act.

Privacy

Drachsler and Greller (2016) defined ethics as a moral code of external conventions in society, while privacy is an intrinsic aspect of a human's identity. Privacy and data protection issues for LA include how personal data is collected and processed by different stakeholders (Ifenthaler & Schumacher, 2016). A broad legal definition of privacy is a human's right to define access to their data and, in the context of learning, to protect a learner's identity to prevent abuse (Dyckhoff et al., 2012). Thus, privacy should be studied as a three-part relationship between a person, some records of information, and other people (Angeli et al., 2017; Manca et al., 2016). More than half of the studied articles mentioned privacy (Fig. 3), which emphasises the importance of the issue/value to develop trust in LA environments (i.e., in the data management layer).

Views on privacy

Privacy is a basic human need; however, a problem with big data is that it is global and permanent. In the past, stakeholders addressed privacy through trust, but some stakeholders do not trust one another in LA. In addition, many institutions do not control the

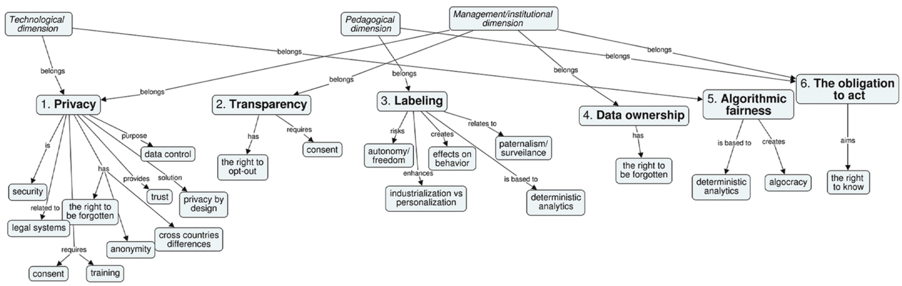

Fig. 2 Concept and relationship mapping of key ethical issues

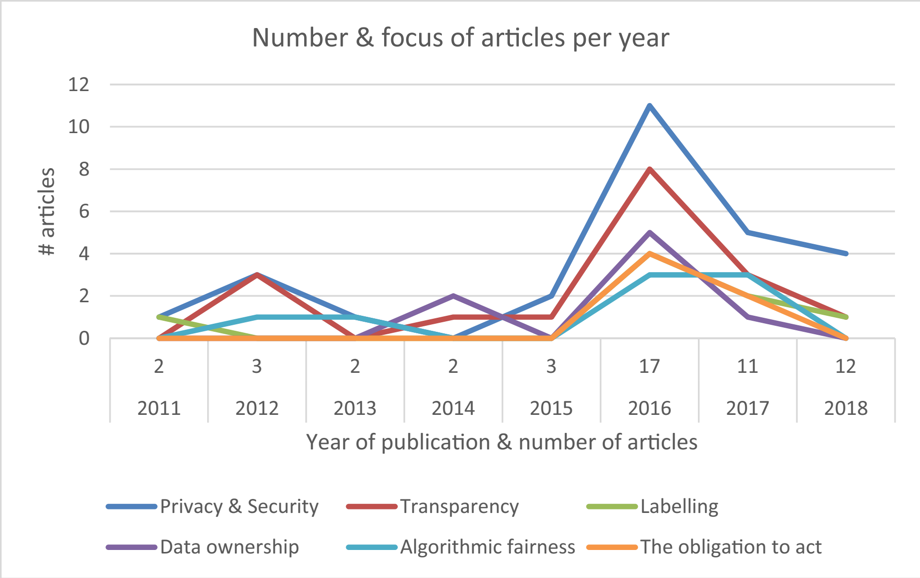

Fig. 3 The focus of articles per year

storage of learners' data because data is handled outside of these institutions or, even worse, outside of the country where these institutions are based, where different privacy laws may apply (Willis et al., 2016). Furthermore, the boundaries and meanings of privacy differ among cultures, and different countries have different ideas about what is ethical.

In the United States (US), no single national law regulates the use of personal data. The US has a system of federal and state regulations that can overlap and contradict one another. Such regulations are US government policies relating to information. They include the Digital Millennium Copyright Act (DMCA), the Family Educational Rights and Protection Act (FERPA), the American Psychology Association (APA), the Health Insurance Portability and Accountability Act (HIPPA), and the Freedom of Information Act (FOIA; Haythornthwaite, 2017). In the US, the collected data belongs to the data collectors, while in the European Union (EU), personal data belongs to the individual who extracts the data.

In the EU, the General Data Protection Regulation (GDPR) (The European Union, 2016) was activated in 2018 to introduce stricter rules on how data collectors may use citizens' information. The GDPR standards relate to informed consent, privacy-by-design, algorithmic transparency, and automated decision-making (Wachter, 2018), so the GDPR provides privacy and control protection for LA users. For example, the consent to use data applies to higher education institutions (HEIs) that offer EU learners courses even if they are outside the EU (Zijlstra-Shaw & Stokes, 2018). Thus, educational institutions must follow these rules in different international settings.

According to Hoel and Chen (2018), the GDPR will influence legislation in countries outside of Europe. Data protection laws (which focus on the individual) have been established in Japan, the Philippines, South Korea, and Taiwan, while in China, Indonesia, and Thailand, all learning activity data is available for analysis (i.e., the focus is on the organisation). In Japan, the Act on the Protection of Personal Information has introduced the concept of anonymously processed data, which makes identifying a specific individual challenging.

Finally, according to the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, personal information is a personal asset in Australia. The ethics concerning this issue consequently differ around the world and between institutions (Willis et al., 2016), and in this sense, establishing policies, practices, and standards between entities is essential as data management methods that align the privacy value. For instance, in the practice of privacy-by-design, privacy is embedded, both in technological and legal terms, at all stages of data collection, storing, and feedback (Gursoy et al., 2017).

With LA that works with sensitive datasets containing personal information, security is a primary concern for ensuring data protection (Pardos et al., 2016). If one of the problems with data protection is to eliminate identifiers, then the information that might lead to cross-checks connected to learners' identification should be hidden. A method is the personal data to be anonymised before the student model can analyse it before presenting the output as recommendations (Cruz et al., 2015; van der Schaaf et al., 2017). Ensuring data anonymity (e.g., noise addition, aggregation, and differential privacy) is critical to data handling, technical tools, organisational frameworks, and the law must develop safeguards (Reidenberg, & Schaub, 2018).

Wintrup (2017) stated that consent and anonymity are necessary when data analytics is combined with characteristics (e.g., age or prior educational achievements). In this regard, Herder and Kawase (2012) stated that anonymisation is often insufficient for securing privacy issues: "how little data it takes to reidentify individuals" (Haythornthwaite, 2017, p. 2). For this reason, there is interest for data analysts in obscuring log data as a method to protect the privacy of individual users while that data remains useful for providing personalised services. An extreme method would be to encrypt the data logs. In this case, any qualitative analysis or interpretation of the data would be impossible. Another privacy-preserving technique is the aggregation of data from many learners; however, analysing individual user behaviour would be challenging; anonymisation implies a compromise between privacy and fidelity (Herder & Kawase, 2012).

In a study by Cruz et al. (2015), the concept of access management describes the procedure of allowing access to protected data. It includes authentication, authorisation, and trust because students share a large amount of personal information, and access to that information should be protected. In contrast, others have suggested that privacy may be voluntarily sacrificed in exchange for learning benefits: "The sharing of personal data with an institution in exchange for better support and personalised learning will be seen as a fair value exchange" (Siemens, 2013, p. 1394).

Transparency

From a data management layer perspective (Gibbons & Rogers, 2009), transparency involves a well-informed choice to opt-in or opt-out. From a pedagogical perspective, this means providing students with self-control and self-observation. It includes information about the following: the parties that have access to data; the data that is collected and visualised; the processing principles (e.g., predictive models and ML algorithms) that are used; and the length of time for which data and outcomes will be stored (Pardo & Siemens, 2014). 18 articles have reported transparency; therefore, it is the second most prevalent issue.

Views on transparency

Consent is ongoing and refers to an individual permitting data gathering and allowing action to be taken based on data processing results. For example, a K-12 school must inform the parents about collecting and analysing their children's data to consent. In the 2014–2015 school year, the National Center for Fair and Open Testing in the US had led one case of opt-out that encouraged approximately 600,000 students to refuse to take government-mandated standardised tests and finally to opt-out. Furthermore, they noted that respecting learners involves allowing them to make their own decisions. In the LA context, this requires learners' participation to result from an informed, voluntary choice. Institutional review boards (IRBs) must approve informed consent forms that learners must sign before they can participate in LA designs, and learners who decline to participate will not be disadvantaged because of their decision (Herder & Kawase, 2012).

Arnold and Sclater (2017) and Hoel et al. (2017) agreed that educational institutions' ethical duty is to obtain the best quality educational data to ensure that they provide the best support. The implication is that if learners have the right to opt-out or game the system, this could be unethical because opting out may leave significant gaps in the data set and reduce the efficiency of LA systems for other learners.

In their study, Herder and Kawase (2012) observed that knowledge and confidentiality are the fundamental prerequisites for learners to consent to collect their data. To motivate learners to relinquish their data to be controlled and invest their time and energy into LA projects, an institution must convince them that the project in question is innovative and fair. According to Ifenthaler and Schumacher (2016), the learners' probability of unveiling required information is higher if they expect the advantages to overcome the risk. More specifically, students agree to share their course enrolment data, learning strategy-related test results, and motivation test results for LA purposes. In contrast, students are not willing to share their medical data and data from social media.

Greller and Drachsler (2012) claimed that with the continuous increase in sensors, the Internet of Things (IoT), and other innovative technologies, many aspects of individual behaviour are being recorded without the data subject's awareness. Moreover, to handle negative behaviour and complaints about LA research, institutions should identify a person who will receive and handle these complaints (Hoel et al., 2017).

In an evidence-based study, West et al. (2016) presented the findings of an Australian study that comprised two surveys: one regarding institutional leaders ($n=22$) and one regarding academic staff ($n=353$). The findings demonstrated a lack of understanding and awareness of ethical LA issues among the key players. In another work (Ifenthaler & Schumacher, 2016), 330 university students participated in an exploratory study. The study's findings indicated that students were conservative in sharing personal data and that learners would share more data if the LA task transparently presented meaningful information.

Labelling

From a pedagogical perspective, it is crucial to allow students to make mistakes and learn from past experiences without their student profile being "etched like a tattoo into their digital skins" (Mayer-Schonberger, 2011, p. 14). Academics sometimes interpret a student's individualised historical data to categorise them based on their estimated degree of success (Lawson et al., 2016). However, analysis based on individual students' characteristics at the start of their study must not guide stakeholders to limit the learning expectations of the university or students (Gray & Boling, 2016; Scholes, 2016). Ten articles reported labelling as the third most frequent issue/value. The selected articles extracted matters of opinion that the following section discusses.

Views on labelling

Student success is a multidimensional phenomenon, and statistics-based decision-making does not assess learners as individuals (Scholes, 2016). Although data-driven education has the advantage of improving learning and increasing educational retention, it could lead to students' labelling—the concern that students may be stereotyped and mistreated (Scholes, 2016). Therefore, the instructors must guarantee that their feedback does not discourage or manipulate students, considering that analytics is not always the outcome of a set of independent variables. Every LA intervention should be applied by following a specific instructional design theory, such as self-regulated learning (SRL) (Pardo et al., 2017; Zimmerman, 1990), in the strategy layer (Gibbons & Rogers, 2009), defining the focus of feedback to have positive influences and providing motivation to students (Wong et al., 2019). For instance, a return to behaviourism as a learning theory would not be the most pedagogically appropriate because a misdirected intervention could result in student effects on behaviour (Siemens & Long, 2011).

Beattie, Woodley, and Souter (2008) argued that learning behaviour predictions are probabilistic and not deterministic. Furthermore, Wintrup (2017) stated that knowledge of student profiles hides the risks that students might be categorised collectively or individually in negative ways. In a scenario of discrimination, an institution's success strategy involves improving admissions and excluding poor quality students. Thus, analytics provides a black box that determines who will fail before students have even begun (Beattie et al., 2008), ignoring characteristics that influence learning, such as special needs or multiple intelligences (Reigeluth, 2015).

Jones (2017) presented cases of paternalism as a moral concept in LA activities. With regard to the socio-critical perspective of LA, the author concluded that paternalism is a contentious issue for HEI under certain circumstances, such as in cases of negative freedom. This is because data analytics, as a surveillance technology embedded in LA tools, creates truth-by-data automated decision-making that reduces personal subjectivity. Although an institution may have positive intentions, surveillance results reduce students' autonomy and risk their academic freedom.

Data ownership

Data ownership is a complicated legal and moral issue in the data management layer. Raw data traces belong to the data subject; however, in practice, processed data no longer belongs to the learner. Ownership refers to the data collected, the analytics used, and the analytics' output, and ten articles reported it.

Views on data ownership

Pardo and Siemens (2014) asked the following questions: (1) Can students control how their data is used and shared? (2) Who owns the data—the institutions as the providers of the technical infrastructure, the students, or the companies that use the data to create educational products? (3) Even if one accepts that the raw data collected from users belongs to them, what happens with the analysed data derived from that raw data? For example, a core part of LA techniques relies on combining students' data to produce a prognosis model. The question that then arises is, who is the owner of such a model? Users have provided the raw material, but a third party implemented the process of creating such a model (Ari & Brandon, 2014; Pardo & Siemens, 2014).

Hoel et al. (2017) referred to learners' "right to be forgotten," which relates to the minimisation of data and the limitation of its use. This right consequently focuses on purposeful data collection and enhances information overload control, as more learning data does not ever make better educational data (Ifenthaler & Tracey, 2016). Students can force institutions to erase data that serve no current purpose through the right to be forgotten. In summary, data ownership usually overlaps with other values about data management goals (see Table 5).

Algorithmic fairness

There are many reasons for the existence of errors in data analysis, including the misinterpretation of data (human error) and the adherence to misleading patterns (machine-based error) (Fynn, 2016). Systems are only as sound as the data; this means that incomplete, noisy, and unrepresentative data or incorrect models lead to misinformed decisions. Standard statistical techniques are inaccurate when they are applied to unstructured textual data. As a consequence, from a pedagogical perspective, the results can be harmful to learners. Seven articles have reported this issue.

Views on algorithmic fairness

Six decades ago, W. Cameron (1958, p. 173) stated that "not everything that can be counted counts; and not everything that counts can be counted." Training data is an important issue when one needs to make predictions that will work with future learners. If the future population's characteristics and behaviour differ significantly from those of the present one, such training will have been conducted on past datasets to produce doubtful results (Beattie et al., 2008). ML algorithms estimate parameters to fit a predetermined model, but that model may not be appropriate. Interestingly, if we have huge sparse vectors for training data, we may end up getting poor models or even overfitting the data due to the curse of dimensionality. Sometimes, instructors use a simplistic and highly domain-specific model with errors or unreliable statistical methods because to do so is convenient, and even if they have aggregated data, they analyse the results for a group. Based on this analysis, they present results for the individual (Arnold & Sclater, 2017). However, computers cannot have bias or stereotypes, so they are expected to perform an objective data-driven analysis. Nevertheless, the models, the training data, and the results of these computers are set by humans, whose biases may influence analysis results (Sarkar, 2019). For instance, wealthy schools typically have computerised education, so the data and insights extracted from LA may not accurately reflect the general population. In an evidence-based study, Gursoy et al. (2017) presented the trade-off between students' privacy and LA interventions' utility, which is affected by the LA tasks' accuracy and performance. Based on the results, the authors stated a matter of opinion that increasing the level of privacy reduces the accuracy of the LA outcomes.

Arnold and Sclater (2017) argued that another fear among instructors is losing students' autonomy and violating learning-focused principles. In a relevant study, Beattie et al. (2008) concluded that with predictive analytics, instructors could irresponsibly profile students on their future performance, announcing who is going to fail a priori. Furthermore, Timmis et al. (2016) stated that technological determinism views technology as the only factor in predetermined changes. This matter of opinion hides the real risk of data-driven institutes developing based on standardised instruction and the student's mitigation to a simple metric (Arnold & Sclater, 2017).

Siemens (2013) and Fynn (2016) stated that the learning process is social and cannot be reduced entirely to algorithms. After all, social constructivism and SRL strategies propose that LA should focus on learners' interactions with their instructors, the learning content, and other learners. The learning process is creative and requires the generation of new ideas. In contrast, analytics involves identifying and presenting that which already exists, leading to competition between innovation (generating new ideas) and analytics (evaluating what exists in data). In this sense, analytics is perceived as a mechanism for regulating and correcting behaviours (Drachsler & Greller, 2016).

The obligation to act

Students' costs are high in terms of the fees, time, and energy they spend on their studies. Therefore, from a management and pedagogical perspective, institutions should strive to support and encourage students (Scholes, 2016). Only three articles have reported this matter of opinion, creating a research gap in the literature.

Views

It is unethical to ignore the predictive value of performance management (West et al., 2016). Prinsloo and Slade (2017) argued that educational stakeholders are ethically responsible for acting when instructional data obligates action. Knowing more about students, monitoring them, and making this knowledge available to stakeholders does not necessarily result in action. This subsection discusses the opinion that HEIs cannot afford not to use data purposefully.

It is unethical not to inform students about their progress and let them continue on a path of academic failure. This matter of opinion is the obligation of knowing, and institutions have the responsibility to implement it through LA, providing timely support after diagnosing outliers (Prinsloo & Slade, 2017).

In addition, students have a co-responsibility to do their best to succeed. According to SRL, students are active users who follow strategies (e.g., time management, goal planning, and self-intervention) (Wong et al., 2019). However, in some cases, students do not share the data they create on an LA system. This paradox may limit the capabilities of LA systems (Ifenthaler & Schumacher, 2016). Applying a similar philosophy, Arnold and Sclater (2017) suggested that students share learning data if the LA system provided meaningful intervention.

Guidelines and summary of findings (RQ1)

Guidelines

Following the previously mentioned organisational scheme, Table 2 summarises the guidelines that correspond to the values about data management goals of privacy, transparency, labelling, ownership, algorithmic fairness, and duty to act.

Transparency and labelling (autonomy)

Beattie et al. (2008) supported the view that learner-oriented LA provides a handy tool that learners own and self-regulate transparently. Furthermore, HEIs should not use data to stereotype learners or other stakeholders negatively. Fundamental principles that determine data management are as follows: student data belongs to the student, should never be shared without informed consent, and improve learning outcomes. In addition, Ari and Brandon (2014) proposed the following principles: (a) LA is a moral practice and should focus on understanding rather than measuring; (b) student profiles and performance are dynamic issues.

Transparency and obligation (duty) to act

First, according to Hoel et al. (2017), the Norwegian Centre for ICT in Education published a guide on LA stipulating that the implementation of LA should follow several principles of data protection: lawfulness, purpose limitation, data minimisation, and the right to know. Second, Slade and Prinsloo (2013) proposed using an ethical framework in which LA should focus on the moral necessity to use information, and students must collaborate with the institution with transparency. Third, Willis et al. (2013) discussed the following ethical principles: HEI can use analytics to identify extreme student behaviours (e.g., under-confidence), institutions must act, and they should use analytics to ensure the success of large numbers of learners.

Algorithmic fairness and labelling

According to Slade (2016), the fundamental principles of LA ethics are as follows: LA is an ethical practice, and the institution has a responsibility to all of its stakeholders. Furthermore, data interpretation should not define students, who should be engaged as active agents, and interventions should be free of bias.

Data ownership and privacy

In a review study, Avella et al. (2016) first addresses issues related to data collection, analysis, and ethics. Informed consent, purposeful data usage, and accountability were the proposed guiding principles. They then discussed mechanisms for transparency and data security.

In another review by Tsai and Gasevic (2017), the authors assessed 25 empirical studies, 38 desk studies, and eight LA policies. The results revealed that stakeholders need more dynamic communication channels in parallel with new pedagogy-based approaches to LA. With regard to challenges, the authors observed a shortage of leadership, pedagogy-based strategies, sufficient training, and LA-specific policies.

In addition, the Learning Analytics Community Exchange (LACE) review (Cardinali et al., 2015) presented policies for educational data mining (using data to make predictions) and LA (using data to make changes that improve learning). The authors focused on the importance of democratic control (i.e., informed consent and ownership); a school's data must not be sold, and analytics should not be the only source of decision-makers' prediction.

A second LACE review (Griffiths et al., 2016) presented cases from different educational sectors in a complex cloud-based environment. The reviewers selected the InBloom case about schools (Arnold & Sclater, 2017), in which foundations, the government, and school managers did not co-operate with other stakeholders, such as parents, who were concerned about privacy issues and data misuse. The weakness in demonstrating its benefits and stakeholders' lack of trust resulted in the termination of this costly LA project after 3 years.

Summary of findings

Our research's final step extracted tables and figures to demonstrate the included studies and their findings in a systematic and definite format. The following quantitative outcomes (Fig. 3) determine the number and focus of articles per year and the number of articles that analysed each issue/value about data management goals.

After studying the literature, the authors chose a summary of findings table τo provide evidence about the types of harm that can accumulate (Table 3). It presents implications and values for various stakeholders, following the opinion that "views on the benefits, risks, and potential for harm resulting from the collection, analysis, and use of student data will depend on the interests and perceptions of the particular stakeholder" (Lang et al., 2018, p. 50). In summary, stress, discrimination, a spoon-feeding learning approach, and effects on behaviour could harm learners, while teachers could make predictions without understanding the model. After all, possible harm articulates the conditions and values behind studying ethics in LA.

After scanning the titles, abstracts, and keywords of the selected 53 papers, a word list, sorted by frequency (Table 4), was extracted to identify the main themes in LA's literature. In summary, learners are the most common target group, and privacy is the most prevalent issue.

In general, this work investigates the overlapping ethical issues (Table 5) that our research identified. It provides insights into the relationships between the values that significantly influence selecting data management methods.

The literature has also addressed many open-ended questions and characteristics of the included studies (Table 6) as useful tools for making complicated issues more manageable.

Finally, Wilson et al. (2017) directed criticism at a lack of pedagogy around big data. Learning analytics as an abstract concept and designed artefact is pedagogically neutral (Greller & Drachsler, 2012). Therefore, a final precondition to addressing LA ethical issues is to define the instruction concepts that HEI must take into account (Ifenthaler, 2017). To do so, we systematically searched the primary sources (562 papers) of our research, studying their abstracts and conclusions. We finally categorised 140 papers, which are empirical articles that contain data-based real case studies that investigate the LA domain in its

Table 3 Stakeholders' ethical issues, responsibilities, and possible harm

Stakeholders	Issues and harm
Learners	Performance-related stress for learners; students' psychological and physical well-being (Reidenberg & Schaub, 2018); profiling based on ML (Peña-Ayala, 2018); spoon-feeding learning approach and risks of demotivation (Tsai et al., 2018); learners being assessed under a microscope; rights of students to remain individuals (Papa & Armfield, 2018); students request to be in a safe environment where they could make mistakes (Drachsler et al., 2015); students to be engaged as collaborators and not as recipients of interventions and services (Slade & Prinsloo, 2013)
Teaching Staff	Teachers accept classification systems as fact despite these processes being subject to data entry errors, data cleaning, and normalisation (Papa & Armfield, 2018); pedagogical expertise needs to be involved in making sense of data and supporting learners to take meaningful action based on the data; confidentiality and design of interventions must be considered (Tsai et al., 2018). There is a power relationship between instructors and students (Lawson et al., 2016)
Institutional actors (instructional designers, administrators)	There is a shortage of leadership to ensure that the implementation of LA is strategically planned (Tsai et al., 2018); the institution does not allow the student to correct data used in the predictive model; predictions are made without understanding the model (Papa & Armfield, 2018); an asymmetrical power relationship exists between institution and student; and outcomes cannot be generalised across institutional and geopolitical contexts (Slade & Prinsloo, 2013). An instructional designer balances the LA community's technocentric tendencies through critical theory and pedagogy (Gray & Boling, 2016)
Vendors (external stakeholders)	Third-party learning environments that track student behaviours present intellectual freedom issues. Furthermore, digital content vendors collect and use data for a variety of reasons, including digital rights management and consumer analytics (Jones & Salo, 2018); vendors rely on data sharing rather than confidentiality (Reidenberg & Schaub, 2018); and there is a lack of adequate technical solutions to ensure opt-out options without affecting the quality of data and services provided (Tsai et al., 2018). Professionals should have an awareness of their ethical role in practice. (Gray & Boling, 2016)

Table 4 Wordlist sorted by frequency (in parentheses)

Stakeholder	Issue	Other
Learners (127)	Privacy (100)	Policy (33)
HE Institutions (88)	Obligation to act (11)	Legal (11)
Teachers (21)	Profiling (10)	IoT (10)
Instructional designers (14)	Transparency (8)	Moral (5)
Librarians (2)	Data ownership (8)	GDPR (5)
Parents (2)	Surveillance (7)	

natural context. We extracted forty-eight studies that mention theories and concepts related to instruction. Table 7 presents the learner-centred instruction paradigms in the strategy layer (Gibbons & Rogers, 2009; Reigeluth & Carr-Chellman, 2009). Consistent with Wong et al. (2019) view, we conclude that SRL ($n = 18$) instructional design theory is the basis

Table 5 The overlapping of ethical issues in the literature

Ethical issues (values about goals)	Number of common articles	Articles
Privacy, transparency	8	Pardos et al. (2016); Willis et al. (2016); Scholes (2016); Herder and Kawase (2012); Greller and Drachsler (2012); Cruz et al. (2015); Ifenthaler and Tracey (2016); Dyckhoff et al. (2012)
Ownership, privacy, transparency	5	West et al. (2016); Avella et al. (2016); Pardo and Siemens (2014); Ifenthaler and Schumacher (2016); Hoel et al. (2017)
Labelling, ownership, privacy	2	Slade and Prinsloo (2013); Ari and Brandon (2014)
Labelling, privacy	2	Wintrup (2017); Siemens and Long (2011)
Labelling, ownership, privacy, transparency	1	Lawson et al. (2016)
Fairness, labelling, privacy	1	Beattie et al. (2008)
Act, fairness, transparency	1	Arnold and Sclater (2017)

for most studies. Self-regulated learning models learners as active users who follow strategies (e.g., time management, self-observation, and goal planning) for success. Self-regulated learning involves, encourages, and motivates learners, who can consequently self-reflect, collaborate, and control their performance. When applied to trace data, this theory explains why students' behaviour varies, and subsequently, personalised feedback, interventions, and scaffolding can be designed (Zimmerman, 1990). In addition, motivational design (ARCS model) (Keller, 2010), engagement, feedback, and active learning are the secondary instruction concepts that LA practices must consider.

A proposed checklist (RQ2)

Thus far, we have described the LA ethics situation with values about goals about instructional data filling out the analysis elements. To address the second research question (what should a proposed checklist for LA ethics include for specific educational stakeholders?), we synthesise and propose our instructional design theory with the methods filling out the solution elements. A checklist with methods is based on the hierarchy of intersecting values about goals in Fig. 2: privacy, autonomy, non-probabilistic algorithms, duty to act, openness and transparency, resolve the data ownership, and all stakeholders (PANDORA). The checklist framework is as follows:

- The main priorities of ethical issues (i.e., the instructional values about goals that drive the decision to use particular methods) and a description per stakeholder.
- The most valued checklist methods can be assessed and marked as present or not, to be a real service.
- The possible harm if the current issue is not faced articulates the needs behind studying LA's ethics.

Table 6 Research and open-ended questions extracted from the literature

Article	Questions/key perspectives
Avella et al. (2016)	What are the challenges of using LA in education?
Pardos et al. (2016)	Transparency: what data is being collected, and how is it being represented?
Greller and Drachsler (2012)	Privacy: Is the analysis following privacy arrangements, and are the students adequately informed?
Pardo and Siemens (2014)	How are privacy and ethics addressed in other contexts? Who owns the data: the institutions, the students, or the companies using them?
Scholes (2016)	Should a decision-maker sort students from group-risk statistics?
Slade and Prinsloo (2013)	Do some labels exist that should be prohibited? Are there circumstances in which other principles override the need for informed consent? Is it ethical to ignore the predictive value of research evidence?
West et al. (2016)	What ethical principles should conduct the use of LA?
Siemens and Long (2011)	If we confine analytics to behavioural data, then how can we account for more than behavioural data?
Sclater (2016)	In which conditions should learners be asked for consent to the collection of their data for analytics?
Siemens (2013)	Who has access to analytics? Should a learner be able to see what an institution sees? How long does a university keep this data?
Hoel et al. (2017)	How will the school ensure that information is used for learning and not for other purposes?
Arnold and Sclater (2017)	Would learners be happy for data on their learning activities to be used if it kept them from dropping out?
Prinsloo and Slade (2017)	How do we respond to the moral and legal necessity to act when responding in appropriate and effective ways becomes impossible?
Drachsler and Greller (2016)	If a computational model is developed from a collection of data traces in a system, can a student still opt-out of such a data model?
Cardinali et al. (2015)	Are there legitimate issues about the impact of data analytics on education?
Griffiths et al. (2016)	What are the ethical implications of knowing, of not knowing, and of refusing to know?
Peña-Ayala (2018)	How is it possible that LA duty co-exists with students without hurting their interests and rights? How does one guarantee that LA labour never puts students' natural daily life at risk?

1. Privacy

1.1 For **institutions or instructional designers,** the institutions should establish security, data management, data minimisation, and control from an administrative perspective. The checklist of data management methods for collecting data ethically for the privacy values about goals is as follows:

1.1.A. Be clear about who has specific access to the recorded data.
1.1.B Develop contracts with externals vendors in ways that respect and manage privacy.
1.1.C. Apply the GDPR.
1.1.D Apply authentication and authorisation techniques.
1.1.E. Hire a data protection officer who will be responsible for compliance with

Table 7 Instructional theory and methods matched from the selected articles

Instructional theories and methods	References
SRL instructional design theory ($n=18$)	Ott et al. (2015); Lu et al. (2017); Pardo et al. (2017); Tabuenca et al. (2015); Park and Jo (2015); Gewerc et al. (2016); Papamitsiou and Economides (2015); Martin and Whitmer (2016); Petropoulou et al. (2014); Stefan et al. (2016); Ruipérez-Valiente et al. (2015); Mazarakis (2014); Chou et al. (2017); Melero et al. (2015); Softic et al. (2014); Olmos and Corrin (2012); Kim et al. (2016); Gasevic et al. (2017)
Engagement instructional outcome ($n=16$)	Kim et al. (2016); Stefan et al. (2016); O'Riordan et al. (2016); Olmos and Corrin (2012); Smith et al. (2012); Tempelaar et al. (2015); Pursel et al. (2016); Davidson and Candy (2016); Lu et al. (2017); Pardo et al. (2017); Ott et al. (2015); Papamitsiou and Economides (2015); Xie et al. (2014); Lan et al. (2014); Sedrakyan et al. (2014); Ma et al. (2014)
Feedback instructional method ($n=14$)	Gibson and de Freitas (2016); Gasevic et al. (2016); Tabuenca et al. (2015); Lan et al. (2014); Chou et al. (2017); Ott et al. (2015); Liu et al. (2016); Kim et al. (2016); Poitras et al. (2016); Firat (2017); Tempelaar et al. (2015), Ifenthaler and Widanapathirana (2014); Kennedy et al. (2013); Lu et al. (2017)
Active learning instructional method ($n=10$)	Gasevic et al. (2016); Mazarakis (2014); Kotsiantis et al. (2014); Petropoulou et al. (2014); Liu et al. (2016); Xie et al. (2014); Gewerc et al. (2016); Xing et al. (2015); Hernández-García et al. (2016); Park and Jo (2015)
Motivational design (ARCS instructional model) ($n=6$)	Tempelaar et al. (2015); Lan et al. (2014); Sedrakyan et al. (2014); Davidson and Candy (2016); Lonn et al. (2015); Mazarakis (2014)

the rules through learning.

1.1.F. Ensure the instructional designer's ethical training and awareness of ethical concerns at all LA process stages.

1.2 For **learners**, consent should be guaranteed from a research perspective; learners should be able to opt-out without adverse consequences, and purposeful LA should be ensured for learners from a legal perspective.

1.2.A. Anonymise students' personal data.
1.2.B. Inform students about the analysis of their learning data.

Without solving this issue with the data management layer, the *harm* is that stakeholders will not trust the LA services and may cancel LA projects.

2. Autonomy

2.1. For **learners and teachers**, the stakeholders' values about goals are to check for intellectual freedom, ensure individuality, and avoid labelling and surveillance. These matters of opinion drive the selection of useful data methods:

2.1.A. Guarantee that the feedback from instructors does not discourage students.
2.1.B. Do not use labels for students that hinder their education and well-being.
2.1.C. Follow a specific instructional design theory (e.g., SRL) to model students as active users.
2.1.D. Respect diverse characters and different learning paths and needs.
2.1.E. Design interpretable recommendation models.

With regard to *harm*, learners fear bias and stigma, and they accept untrusted categorization or unfair decisions; they are consequently passive recipients from a pedagogical perspective. Learners also feel discouraged and that their academic freedom is at risk, both of which limit their learning expectations.

3. Non-probabilistic algorithms

3.1. For **institutions**, the following instructional values about goals should be ensured: the quality and objectivity of data and models, the absence of interventionism, and the utilisation of learner-oriented approaches. The most prominent methods are as follows:

3.1.A. Take into account that a student's performance has a temporal and dynamic character.
3.1.B. Inform data administrators about the processing principles employed (e.g., predictive models, ML algorithms).
3.1.C. Make biases explicit in order to overcome them.
3.1.D. Make use of representative data.
3.1.E. Manage imbalances and inequalities within data sets.

3.2. For **teachers**, the possibility of a human or machine-based error exists, so misdirected interventions should be considered.

3.2.A. Explain to students how the models produce reliable outcomes and why they have been selected for intervention.
3.2.B. Use SRL to trace data and analysis to extract insights into the reasons for variation in students' behaviour.
3.2.C. Take into account that the features in a predictive model are usually limited in accordance with the training vector space.
3.2.D. Try to understand the reason for misclassifications or wrong predictions because no model is 100% accurate.

3.3 For **learners**, learning is not a deterministic procedure.

3.3.A. Inform students that LA should not be the only source of decision-making.
3.3.B. Train learners to interpret the results and visualisations of LA critically.

Without addressing this issue, the *harm* for learners is that they will lose their autonomy. Institutions make predictions without understanding the model, thereby reducing the LA outcomes' accuracy and creating biases in data interpretation.

4. Duty to act

 4.1. For **learners**, the right to know should be applied as a moral (matter of opinion) and legal (matter of fact) necessity to act.

  4.1.A. Inform students about their progress and provide timely support.
  4.1.B. Encourage self-interventions for learners.

 4.2. For **teachers and institutions**, accurate and timely interventions should be provided.

  4.2.A. Take into account the predictive value of LA.
  4.2.B. Connect the specific types of data embedded in LA with specific instructional design theories (e.g., SRL with time spent using learning management systems).
  4.2.C. Use early alert systems to achieve positive student motivation.
  4.2.D. Do not ignore ethics (e.g., follow a guideline).
  4.2.E. Inform instructional designers if the intervention is more harmful than beneficial to the welfare of the learner.

When stakeholders violate the above-mentioned methods, first, the *harm* for learners is that timely support is not provided. Second, communication and trust among stakeholders decrease. Finally, it is costly for students to study and withdraw from education in fees, time, and energy.

5. Openness and transparency

 5.1. For **learners**, the possibility for informed and voluntary consent should be provided. This value drives the below data management methods:

  5.1.A. A student can see what an institution sees.
  5.1.B. A student can opt-out of (or not opt-in to) a data model.
  5.1.C. Students' data should never be shared without their informed consent.
  5.1.D. The institution must appoint a person to handle complaints about LA research.

 5.2. For **institutions**, purpose limitation should be imposed, and their awareness of data use and algorithms are matters of opinion that should be ensured.

  5.2.A. Ensure that student data will not be sold.
  5.2.B. Ensure that information is used for learning and not for other purposes.
  5.2.C. Define the data that is being collected, why and how it is being collected and visualised.
  5.2.D. Define who has accountability for the overall LA procedure.
  5.2.E. Encourage academics to use the LA system in a manner consistent with the

course designers' intentions.

If we ignore this issue/value, learners will become stressed and demotivated to provide their analysis data.

6. Resolve the data ownership

   6.1. **Learners** must have the right to be forgotten. Following a user-centric design aims to place students in control of their data.

      6.1.A. Define the duration for which data and outcomes will be stored.
      6.1.B. Students have the right to correct inaccurate information and remove irrelevant information.
      6.1.C. Students can control how their data is used and shared.

   6.2. **Institutions** should take on the responsibility and control of data and data processing.

      6.2.A. Issue data access specific permissions to each stakeholder.
      6.2.B. Take into account the different laws between countries and the different approaches among institutions.
      6.2.C. Handle information about the learners securely.

If stakeholders do not resolve this issue, the *harm* is that learners will not trust the LA services and will hide their learning data.

7. **All stakeholders** (i.e., students, instructors, institutions, and industrial agents) should be involved and communicate with each other (Table 3). The methods for this issue are as follows:

7.1.A. Inform learners about their responsibility for self-intervention.
7.1.B. Provide teachers and data administrators with sufficient training in LA.
7.1.C. Establish channels of communication between stakeholders (e.g., IRBs or parents as partners in learning).
7.1.D. Establish data ethics teams within institutions with experts in data ethics and representatives of faculties and students.
7.1.E. Train educational technology staff in analytical skills (e.g., in using algorithms and statistics to design and implement LA initiatives).
7.1.F. Ensure that LA stakeholders and interdisciplinary practitioners (e.g., teachers and librarians) have professional codes of ethics (e.g., library ethics).
7.1.G. Ground the designs of LA models in learning theories.

Instructional designers are responsible for designing and implementing an LA system, but they should also ensure that this LA system will be a safe learning experience. The key agents should feel that LA ethics offers them more purposeful benefits than harm. Without considering this issue, the *harm* is that the stakeholders will have no responsibility or means of communication. Moreover, students will be engaged as recipients of (and not as collaborators in) interventions and LA services. Thus, overall, an asymmetrical power relationship will exist between data gatherers and the data object.

Discussion

Classification scheme

Instructional design theory and the classification scheme with the values about goals (Fig. 2) of LA ethics are used as heuristic tools and an organising framework for the proposed PANDORA checklist. We emphasise that applying these rules/methods should not be mechanical but should instead require the key players' awareness and interaction. This LA ethics instructional design encourages stakeholders to be ethically prepared, and it offers clarity and acceptability of LA tasks in the data management layer. It covers the terms of the trust, access, and accountability from the viewpoints of policy, technology, practice, and legislation. The proposed checklist for the LA ethics application can be credibly implemented if all critical dimensions/values about goals are considered. After all, the PANDORA checklist follows Lang et al. (2018)'s suggestion that since LA applications may be developed outside universities and governments by private companies, establishing a code of ethics is essential.

Our opinion is that institutions should strike a balance between, on the one hand, protecting learners' privacy to establish trust and, on the other hand, collecting all necessary data to achieve their purpose to support teaching and learning on a personalised basis. Moreover, the rights to privacy, autonomy, and consent are not independent; stakeholders should consider them associated with other rights (e.g., the right to know, refuse to know, and do one's best).

Without transparency in the LA process, learners will face fear and resistance instead of the desired trust. Furthermore, learners should be allowed to make mistakes and learn from them without the sensation of surveillance and a fear of the consequences. Learners must be aware of what is taking place; they should have the choice to decide for themselves and ask for pedagogical support in an environment that enhances the acceptance of divergence.

Key stakeholders should know what is occurring inside the black box of algorithms to gain an in-depth understanding of the nature, objectives, and boundaries of data-driven decision-making in co-operation with vendors of LA technology. In addition, students want to know how the LA models produce reliable outcomes transparently and why they have been selected for intervention. Institutions must, therefore, establish cross-stakeholder communication strategies to explain how individuals are affected by complicated LA tools, engaging stakeholders in peer critique.

Similarly, Drachsler and Greller (2016) proposed a checklist with the acronym DELICATE to establish appropriate LA; LACE has recommended it, and the LA community uses it as an applicable state-of-the-art instrument for any educational institution. The authors summarise these principles as follows:

- Determination: Decide the added value of LA.
- Explain: Define the scope of data collection.
- Legitimate: Establish how to operate within legal frameworks.

- Involve: Talk to stakeholders.
- Consent: Establish informed consent.
- Anonymise: De-identify individuals.
- Technical aspects: Ensure privacy or security.
- External partners: Ensure that external partners abide by the contract.

After comparing the PANDORA and DELICATE checklists, which enable LA to be ethical, we have identified the below similarities and differences:

- The “privacy” issue from PANDORA shares similarities with the following issues from the DELICATE checklist: “anonymise”, “technical aspects”, and “external partners”.
- The “openness and transparency” issue shares similarities with “explain” and “consent”.
- “Resolve data ownership” is similar to “technical aspects” and “determination”.
- “All stakeholders” communication shares similarities with “external partners” and “involve”.
- The DELICATE checklist does not include the following issues: “autonomy”, “non-probabilistic algorithms”, and “duty to act”.
- The PANDORA checklist does not mention the “legitimate” DELICATE issue.

Furthermore, PANDORA is founded on a rule-based consequentialist ethical position to be adaptive and scalable. PANDORA is an instructional design of LA that selects a taxonomy of methods and principles that enable LA to be ethically based on specific values about goals.

Conclusion and research opportunities

Ethics is a factor that mediates the adoption and impact of LA. This review aimed to address the following questions “what is essential in LA ethics for key educational stakeholders, and what should a proposed checklist for LA ethics include for specific educational stakeholders?” It documents the ethical concerns related to LA and the management of big data associated with education. The inclusion of instructional theory and the Gibbons and Rogers (2009) “layering” framework contributed to framing this research under a consistent theoretical hierarchy. This hierarchy starts with a situation (LA ethics), then extracts conditions (matters of fact), specifically, instruction development constraints and values (matters of opinion), specifically, values about goals (Fig. 2). On the third level, these conditions and values influence the selection of methods that reflect the data management layer across all stakeholders. Finally, the proposed instructional design theory constructs a useful framework for (1) reflecting LA ethics to a theoretical foundation to which LA researchers could evolve and adapt to a changing field, and it would help LA practitioners reclaim, and (2) showing how values about goals lead to the recommendation of important data methods.

The conclusion is that PANDORA contributes as a starting point for a dialogue on rethinking LA ethics and building an agenda of credible solutions. The results indicate that ethical analysis is a pedagogical precondition for successful, robust LA environments. Furthermore, we note that the LA research community is focusing on LA ethics each year (see distribution in Fig. 3) with a significant increase in the number of published articles, in accordance with another review (Viberg et al., 2018).

Overall, the reviewed papers extract the following promising research directions and opportunities. Concerning privacy, data ownership, and transparency, with the continuous increase of innovative technologies (e.g., drones, GPS student tracking systems, sensors, IoT, and face recognition systems), many individual behaviour aspects are recorded without the data subject's awareness and consent. Further research is needed to implement the technology to allow learners to monitor and manage an institution's access to their data.

Concerning stakeholders and governance, an open research question asks how the LA ethics policies from different countries will converge for the broader benefit of education, while the educational systems in different countries have fundamental differences in their philosophies and infrastructures.

Concerning the duty to act (the right to know), only three articles have reported this issue; therefore, a research gap exists in the literature. More empirical research is required to specify the conditions (e.g., the ethical framework) under which learners would be motivated to share their data lifecycles.

The findings demonstrate the shortage of empirically evident guidelines on LA ethics. To generalise the design, more case studies are required that describe current practices and experiences in the use of data analytics ethics in higher education. In addition, to obtain sound evidence, further research is needed from academia, practitioners, and industry to determine the intersection of theory, data, and practice. Future work should apply the proposed ethics-by-design PANDORA checklist as a real service in real educational conditions to improve the methods' ability to achieve the ethical goals under given conditions and values. We acknowledge that future research must be conducted to verify these data management methods in different educational learner-centred settings for various stakeholders and cultures. Finally, we intend to validate our proposed checklist with expert feedback and field testing.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Angeli, C., Howard, S. K., Ma, J., Yang, J., & Kirschner, P. A. (2017). Data mining in educational technology classroom research: Can it make a contribution? *Computers & Education, 113,* 226–242. https://doi.org/10.1016/j.compedu.2017.05.021

Ari, J., & Brandon, L. (2014). *Learning analytics, From research to practice.* Springer, (Vol. 55). Retrieved from http://ci.nii.ac.jp/naid/110009751931/

Arnold, K. E., & Sclater, N. (2017). Student perceptions of their privacy in learning analytics applications. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on—LAK'17,* 66–69.https://doi.org/10.1145/3027385.3027392

Asamoah, D. A., Sharda, R., Hassan Zadeh, A., & Kalgotra, P. (2017). Preparing a data scientist: A pedagogic experience in designing a big data analytics course. *Decision Sciences Journal of Innovative Education, 15*(2), 161–190. https://doi.org/10.1111/dsji.12125

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2), 13–29. https://doi.org/10.24059/olj.v20i2.790

Beattie, S., Woodley, C., & Souter, K. (2008). Creepy analytics and learner data rights. *Ascilite.Org*, 421–425. Retrieved from http://ascilite.org/conferences/dunedin2014/files/concisepapers/69-Beattie.pdf
Cameron, W. B. (1958). *Tell me not in mournful numbers*. National Education Association of the United States.
Cardinali, F., Ferguson, R., Griffiths, D., Hoel, T., Karlberg, P., & Paini, M. (2015). Policy recommendations for learning analytics from three stakeholder workshops- *Learning Analytics Review. July, 10 (August),* 1–6. Retrieved from http://www.laceproject.eu/publications/policy-recommendations-lace-workshops.pdf
Chou, C. Y., Tseng, S. F., Chih, W. C., Chen, Z. H., Chao, P. Y., Lai, K. R.,…Lin, Y. L. (2017). Open student models of core competencies at the curriculum level: Using learning analytics for student reflection. *IEEE Transactions on Emerging Topics in Computing*, *5*(1), 32–44.https://doi.org/10.1109/TETC.2015.2501805
Cruz, G., Costa, A., Martins, P., Goncalves, R., & Barroso, J. (2015). Toward educational virtual worlds: Should identity federation be a concern? *Journal of Educational Technology & Society, 18,* 27–36.
Davidson, S. J., & Candy, L. (2016). Teaching EBP using game-based learning: Improving the student experience. *Worldviews on Evidence-Based Nursing, 13*(4), 285–293. https://doi.org/10.1111/wvn.12152
Drachsler, H., Cooper, A., Hoel, T., Ferguson, R., Berg, A., Scheffel, M.,…Chen, W. (2015). Ethical and privacy issues in the application of learning analytics. *5th International Learning Analytics & Knowledge Conference (LAK15).* https://doi.org/10.1145/2723576.2723642
Drachsler, H., & Greller, W. (2016). Privacy and analytics—It' s a DELICATE issue A checklist for trusted learning analytics. *6th Conference on Learning Analytics and Knowledge*. https://doi.org/10.1145/2883851.2883893
Dyckhoff, A. L., Sielke, D., Bultman, M., Chatti, M. A., & Schroeder, U. (2012). Design and implementation of a learning analytics toolkit for teachers. *Journal of Educational Technology & Society, 15*(3), 58–76.
Firat, M. (2017). Determining the effects of LMS learning behaviors on academic achievement in a learning analytic perspective. *Journal of Information Technology Education Research, 15,* 075–087. https://doi.org/10.28945/3405
Fynn, A. (2016). Ethical considerations in the practical application of the Unisa socio-critical model of student success. *International Review of Research in Open and Distance Learning, 17*(6), 206–220.
Gasevic, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *Internet and Higher Education, 28,* 68–84. https://doi.org/10.1016/j.iheduc.2015.10.002
Gasevic, D., Mirriahi, N., Dawson, S., & Joksimovic, S. (2017). Effects of instructional conditions and experience on the adoption of a learning tool. *Computers in Human Behavior, 67,* 207–220. https://doi.org/10.1016/j.chb.2016.10.026
Gewerc, A., Rodriguez-Groba, A., & Martinez-Pineiro, E. (2016). Academic social networks and learning analytics to explore self-regulated learning: A case study. *Revista Iberoamericana de Tecnologias del Aprendizaje, 11*(3), 159–166. https://doi.org/10.1109/RITA.2016.2589483
Gibbons, A. (2014). *An architectural approach to instructional design*. Routledge. https://doi.org/10.4324/9780203075203
Gibbons, A. S., & Rogers, P. C. (2009). The architecture of instructional theory. In C. M. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base* (Vol. III, pp. 305–326). Lawrence Erlbaum Associates.
Gibson, D., & de Freitas, S. (2016). Exploratory analysis in learning analytics. *Technology, Knowledge and Learning, 21*(1), 5–19. https://doi.org/10.1007/s10758-015-9249-5
Gray, C. M., & Boling, E. (2016). Inscribing ethics and values in designs for learning: A problematic. *Educational Technology Research and Development, 64*(5), 969–1001. https://doi.org/10.1007/s11423-016-9478-x
Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42–57.
Griffiths, D., Drachsler, H., Kickmeier-Rust, M., Steiner, C., Hoel, T., & Greller, W. (2016). Is privacy a show-stopper for learning analytics? A review of current issues and solutions, (February), 1–30. Retrieved from http://www.laceproject.eu/learning-analytics-review/files/2016/04/LACE-review-6_privacy-show-stopper.pdf
Gursoy, M. E., Inan, A., Nergiz, M. E., & Saygin, Y. (2017). Privacy-preserving learning analytics: Challenges and techniques, (114), 1–14.
Haythornthwaite, C. (2017). An information policy perspective on learning analytics. Proceedings of the Seventh International Learning Analytics & Knowledge Conference on—LAK'17, 253–256.https://doi.org/10.1145/3027385.3027389

Herder, E., & Kawase, R. (2012). Considerations for recruiting contributions to anonymised data sets. *International Journal of Technology Enhanced Learning, 4*(1), 85–98. https://doi.org/10.1504/IJTEL.2012.048315
Hernández-García, Á., González-González, I., Jiménez Zarco, A. I., & Chaparro-Peláez, J. (2016). Visualizations of online course interactions for social network learning analytics. *International Journal of Emerging Technologies in Learning, 11*(7), 6–15. https://doi.org/10.3991/ijet.v11i07.5889
Hoel, T., & Chen, W. (2018). Privacy and data protection in learning analytics should be motivated by an educational maxim towards a proposal. *Research and Practice in Technology Enhanced Learning, 13*(1), 20. https://doi.org/10.1186/s41039-018-0086-8
Hoel, T., Griffiths, D., & Chen, W. (2017). The influence of data protection and privacy frameworks on the design of learning analytics systems. Proceedings of the Seventh International Learning Analytics & Knowledge Conference on—LAK'17, 243–252.https://doi.org/10.1145/3027385.3027414
Ifenthaler, D. (2017). Designing effective digital learning environments: Toward learning analytics design. *Technology, Knowledge and Learning, 22*(3), 401–404. https://doi.org/10.1007/s10758-017-9333-0
Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development, 64*(5), 923–938. https://doi.org/10.1007/s11423-016-9477-y
Ifenthaler, D., & Tracey, M. W. (2016). Exploring the relationship of ethics and privacy in learning analytics and design: Implications for the field of educational technology. *Educational Technology Research and Development, 64*(5), 877–880. https://doi.org/10.1007/s11423-016-9480-3
Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1–2), 221–240. https://doi.org/10.1007/s10758-014-9226-4
Jantti, M., & Heath, J. (2016). What role for libraries in learning analytics? *Performance Measurement and Metrics, 17*(2), 203–210. https://doi.org/10.1108/PMM-04-2016-0020
Jones, K. (2017). Learning analytics and its paternalistic influences. *Proceedings from HCI International*. https://doi.org/10.1007/978-3-319-58515-4_22
Jones, K., & Salo, D. (2018). Learning analytics and the academic library: Professional ethics commitments at a crossroads. *College & Research Libraries, 79*(3), 304–323. https://doi.org/10.5860/crl.79.3.304
Keller, J. M. (2010). *Motivational design for learning and performance: The ARCS model approach*. Springer.
Kennedy, G., Ioannou, I., Zhou, Y., Bailey, J., & Oeary, S. (2013). Mining interactions in immersive learning environments for real-time student feedback. *Australasian Journal of Educational Technology, 29*(2), 172–183. https://doi.org/10.14742/ajet.700
Kim, D., Park, Y., Yoon, M., & Jo, I. H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *Internet and Higher Education, 30,* 30–43. https://doi.org/10.1016/j.iheduc.2016.03.002
Kitchenham, B. A. (2004). *Procedures for undertaking systematic reviews, Joint technical report*. Computer Science Department, Keele University (TR/SE- 0401) and National ICT Australia Ltd. (0400011T.1).
Kotsiantis, S., Tselios, N., Filippidi, A., & Komis, V. (2014). Using learning analytics to identify successful learners in a blended learning course. *International Journal of Technology Enhanced Learning, 5*(2), 133. https://doi.org/10.1504/ijtel.2013.059088
Lan, A. S., Studer, C., Waters, A. E., & Baraniuk, R. G. (2014). Tag-aware ordinal sparse factor analysis for learning and content analytics. (March 2013). Retrieved from http://arxiv.org/abs/1412.5967
Lang, C., Macfadyen, L. P., Slade, S., Prinsloo, P., & Sclater, N. (2018). The complexities of developing a personal code of ethics for learning analytics practitioners. https://doi.org/10.1145/3170358.3170396.
Lang, C., Siemens, G., Wise, A., & Gasevic, D. (2017). *Handbook of Learning analytics*. Society for Learning Analytics Research.
Lawson, C., Beer, C., Rossi, D., Moore, T., & Fleming, J. (2016). Identification of "at risk" students using learning analytics: The ethical dilemmas of intervention strategies in a higher education institution. *Educational Technology Research and Development, 64*(5), 957–968. https://doi.org/10.1007/s11423-016-9459-0
Leitner, P., Khalil, M., & Ebner, M. (2017). Learning analytics in higher education—A literature review. In A. Peña-Ayala (Ed.), *Learning analytics: Fundaments, applications, and trends* (pp. 1–23). Springer. https://doi.org/10.1007/978-3-319-52977-6_1
Lin, L., & Spector, J. M. (Eds.). (2017). *The sciences of learning andinstructional design: Constructive articulation betweencommunities*. Routledge. https://doi.org/10.4324/9781315684444

Liu, S., Hu, Z., Peng, X., Liu, Z., Cheng, H. N. H., & Sun, J. (2016). Mining learning behavioral patterns of students by sequence analysis in cloud classroom. *International Journal of Distance Education Technologies, 15*(1), 15–27. https://doi.org/10.4018/ijdet.2017010102
Lonn, S., Aguilar, S. J., & Teasley, S. D. (2015). Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. *Computers in Human Behavior, 47,* 90–97. https://doi.org/10.1016/j.chb.2014.07.013
Lu, O. H. T., Huang, J. C. H., Huang, A. Y. Q., & Yang, S. J. H. (2017). Applying learning analytics for improving students engagement and learning outcomes in a MOOCs enabled collaborative programming course. *Interactive Learning Environments, 25*(2), 220–234. https://doi.org/10.1080/10494820.2016.1278391
Ma, J., Han, X., Yang, J., & Cheng, J. (2014). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *Internet and Higher Education, 24,* 26–34. https://doi.org/10.1016/j.iheduc.2014.09.005
Manca, S., Caviglione, L., & Raffaghelli, J. E. (2016). Big data for social media learning analytics: Potentials and challenges. *Journal of e-Learning and Knowledge Society, 12,* 27–39. https://doi.org/10.20368/1971-8829/1139
Martin, F., & Whitmer, J. C. (2016). Applying learning analytics to investigate timed release in online learning. *Technology, Knowledge and Learning, 21*(1), 59–74. https://doi.org/10.1007/s10758-015-9261-9
Mayer-Schonberger, V. (2011). *Delete: The virtue of forgetting in the digital age*. Princeton University Press. https://doi.org/10.1111/j.1540-5931.2011.921_1.x
Mazarakis, A. (2014). Like diamonds in the sky: How feedback can boost the amount of available data for learning analytics. *International Journal of Technology Enhanced Learning, 5*(2), 107. https://doi.org/10.1504/ijtel.2013.059083
Melero, J., Hernández-Leo, D., Sun, J., Santos, P., & Blat, J. (2015). How was the activity? A visualization support for a case of location-based learning design. *British Journal of Educational Technology, 46*(2), 317–329. https://doi.org/10.1111/bjet.12238
Olmos, M., & Corrin, L. (2012). Learning analytics: A case study of the process of design of visualizations. *Journal of Asynchronous Learning Network, 16*(3), 39–49. https://doi.org/10.24059/olj.v16i3.273
O'Riordan, T., Millard, D. E., & Schulz, J. (2016). How should we measure online learning activity? *Research in Learning Technology, 24*(1063519), 1–15. https://doi.org/10.3402/rlt.v24.30088
Ott, C., Robins, A., Haden, P., & Shephard, K. (2015). Illustrating performance indicators and course characteristics to support students' self-regulated learning in CS1. *Computer Science Education, 25*(2), 174–198. https://doi.org/10.1080/08993408.2015.1033129
Papa, R., & Armfield, S. (2018). The Wiley Handbook of Educational Policy, ISBN:9781119218456. https://doi.org/10.1002/9781119218456.
Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence on JSTOR. *Journal of Educational Technology & Society, 17*(4), 49–64.
Papamitsiou, Z., & Economides, A. (2015). Temporal learning analytics visualizations for increasing awareness during assessment. *RUSC. Universities and Knowledge Society Journal, 12*(3), 129. https://doi.org/10.7238/rusc.v12i3.2519
Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies, 10*(1), 82–92. https://doi.org/10.1109/TLT.2016.2639508
Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology, 45*(3), 438–450. https://doi.org/10.1111/bjet.12152
Pardos, Z., Whyte, A., & Kao, K. (2016). *moocRP: Enabling open learning analytics with an open source platform for data distribution, analysis, and visualization*. Springer.
Park, Y., & Jo, I. H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science, 21*(1), 110–133.
Peña-Ayala, A. (2018). Learning analytics: A glance of evolution, status, and trends according to a proposed taxonomy. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 8*(3), 1–29. https://doi.org/10.1002/widm.1243
Petropoulou, O., Kasimatis, K., Dimopoulos, I., & Retalis, S. (2014). LAe-R: A new learning analytics tool in Moodle for assessing students' performance. *Bulletin of the IEEE Technical Committee on Learning Technology, 16*(1), 1–13.
Poitras, E. G., Naismith, L. M., Doleck, T., & Lajoie, S. P. (2016). Using learning analytics to identify medical student misconceptions in an online virtual patient environment. *Journal of Asynchronous Learning Network*. https://doi.org/10.24059/olj.v20i2.802

Prinsloo, P., & Slade, S. (2017). An elephant in the learning analytics room. Proceedings of the Seventh International Learning Analytics & Knowledge Conference on—LAK'17, 46–55.https://doi.org/10.1145/3027385.3027406

Pursel, B. K., Zhang, L., Jablokow, K. W., Choi, G. W., & Velegol, D. (2016). Understanding MOOC students: Motivations and behaviours indicative of MOOC completion. *Journal of Computer Assisted Learning, 32*(3), 202–217. https://doi.org/10.1111/jcal.12131

Reidenberg, J. R., & Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education, 2017,* 147787851880530. https://doi.org/10.1177/1477878518805308

Reigeluth, C. M. (1999). What is instructional-design theory and how is it changing? In C. Reigeluth (Ed.), *Instructional design theories and models* (Vol. II, pp. 5–30). Lawrence Erlbaum Associates, Inc.

Reigeluth, C. M. (2015). Instructional theory and technology for the new paradigm of education. In: *International handbook of E-learning: Theoretical perspectives and research* (Vol. I) (pp. 79–92). https://doi.org/10.4324/9781315760933.

Reigeluth, C. M., & Carr-Chellman, A. (2009). Understanding instructional theory. In C. M. Reigeluth & A. Carr-Chellman (Eds.), *Instructional-design theories and models: Building a common knowledge base* (Vol. III, pp. 3–26). Lawrence Erlbaum Associates.

Ruipérez-Valiente, J. A., Muñoz-Merino, P. J., Leony, D., & Delgado Kloos, C. (2015). ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy Platform. *Computers in Human Behavior, 47,* 139–148. https://doi.org/10.1016/j.chb.2014.07.002

Sarkar, D. (2019). Text analytics with python. ISBN: 978-1-4842-4353-4. https://doi.org/10.1007/978-1-4842-4354-1

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development, 64*(5), 939–955. https://doi.org/10.1007/s11423-016-9458-1

Sclater, N. (2016). Developing a code of practice for learning analytics. *Journal of Learning Analytics, 3,* 16–42. https://doi.org/10.18608/jla.2016.31.3

Sedrakyan, G., Snoeck, M., & De Weerdt, J. (2014). Process mining analysis of conceptual modeling behavior of novices—Empirical study using JMermaid modeling and experimental logging environment. *Computers in Human Behavior, 41,* 486–503. https://doi.org/10.1016/j.chb.2014.09.054

Siemens, G. (2012). Learning analytics: envisioning a research discipline and a domain of practice. *LAK'12: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*. https://doi.org/10.1145/2330601.2330605

Siemens, G. (2013). Learning analytics. *The American Behavioral Scientist, 57*(10), 1380–1400. https://doi.org/10.1177/0002764213498851

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46,* 30–32. https://doi.org/10.17471/2499-4324/195

Slade, S. (2016). The Open University ethical use of student data for learning analytics policy. https://doi.org/10.13140/RG.2.1.1317.4164.

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. https://doi.org/10.1177/0002764213479366

Smith, V. C., Lange, A., & Huston, D. R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions. *Journal of Asynchronous Learning Networks, 16*(3), 51–61.

Softic, S., De Vocht, L., Taraghi, B., Ebner, M., Mannens, E., & De Walle, R. V. (2014). Leveraging learning analytics in a personal learning environment using linked data. *Bulletin of the Technical Committee on Learning Technology, 16*(4), 10–13.

Spector, J. M. (2015). Foundations of educational technology, ISBN: 978-1-315-76426-9. https://doi.org/10.4324/9781315764269

Spector, J. M. (2016). Ethics in educational technology: Towards a framework for ethical decision making in and for the discipline. *Educational Technology Research and Development, 64*(5), 1003–1011. https://doi.org/10.1007/s11423-016-9483-0

Stefan, L., Moldoveanu, F., & Gheorghiu, D. (2016). Evaluating a mixed-reality 3D virtual campus with big data and learning analytics: A transversal study. *Journal of E-Learning and Knowledge Society, 12*(2), 41–54. https://doi.org/10.20368/1971-8829/1132

Tabuenca, B., Kalz, M., Drachsler, H., & Specht, M. (2015). Time will tell: The role of mobile learning analytics in self-regulated learning. *Computers and Education, 89,* 53–74. https://doi.org/10.1016/j.compedu.2015.08.004

Taylor, M. J., Dove, E. S., Laurie, G., & Townend, D. (2018). When can the child speak for herself? The limits of parental consent in data protection law for health research. *Medical Law Review, 26*(3), 369–391. https://doi.org/10.1093/MEDLAW/FWX052

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47,* 157–167. https://doi.org/10.1016/j.chb.2014.05.038

The European Union. (2016). Regulation (eu) 2016/679 of the European parliament and of the council. *Official Journal of the European Union, L119,* 59.

Timmis, S., Broadfoot, P., Sutherland, R., & Oldfield, A. (2016). Rethinking assessment in a digital age: Opportunities, challenges and risks. *British Educational Research Journal, 42*(3), 454–476. https://doi.org/10.1002/berj.3215

Tsai, Y.-S., & Gasevic, D. (2017). Learning analytics in higher education—Challenges and policies. *Proceedings of the Seventh International Learning Analytics & Knowledge Conference on—LAK'17,* 233–242. https://doi.org/10.1145/3027385.3027400

Tsai, Y.-S., Moreno-Marcos, P. M., Jivet, I., Scheffel, M., Tammets, K., Kollom, K., & Gašević, D. (2018). The SHEILA framework: Informing institutional strategies and policy processes of learning analytics. *Journal of Learning Analytics, 5*(3), 5–20. https://doi.org/10.18608/jla.2018.53.2

van der Schaaf, M., Donkers, J., Slof, B., Moonen-van Loon, J., van Tartwijk, J., Driessen, E.,…Ten Cate, O. (2017). Improving workplace-based assessment and feedback by an E-portfolio enhanced with learning analytics. *Educational Technology Research and Development*, *65*(2), 359–380.https://doi.org/10.1007/s11423-016-9496-8

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89,* 98–110. https://doi.org/10.1016/j.chb.2018.07.027

Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. *Computer Law & Security Review, 34*(3), 436–449. https://doi.org/10.1016/j.clsr.2018.02.002

West, D., Huijser, H., & Heath, D. (2016). Putting an ethical lens on learning analytics. *Educational Technology Research and Development, 64*(5), 903–922. https://doi.org/10.1007/s11423-016-9464-3

Willis, J., Campbell, J., & Pistilli, M. (2013). Ethics, big data and analytics: A model for application. EDUCAUSE Review Online. Retrieved from: http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application.

Willis, J. E., Slade, S., & Prinsloo, P. (2016). Ethical oversight of student data in learning analytics: A typology derived from a cross-continental, cross-institutional perspective. *Educational Technology Research and Development, 64*(5), 881–901. https://doi.org/10.1007/s11423-016-9463-4

Wilson, A., Thompson, T. L., Watson, C., Drew, V., & Doyle, S. (2017). Big data and learning analytics: Singular or plural? *First Monday*. https://doi.org/10.5210/fm.v22i4.6872

Wintrup, J. (2017). Higher education's Panopticon? Learning analytics, ethics and student engagement. *Higher Education Policy, 30*(1), 87–103. https://doi.org/10.1057/s41307-016-0030-8

Wong, J., Baars, M., de Koning, B. B., van der Zee, T., Davis, D., Khalil, M.,...Paas, F. (2019). Educational theories and learning analytics: From data to knowledge. In utilizing learning analytics to support study success (pp. 3–25). Springer.

Xie, C., Zhang, Z., Nourian, S., Pallant, A., & Hazzard, E. (2014). Time series analysis method for assessing engineering design processes using a CAD tool. *International Journal of Engineering Education, 30*(1), 218–230.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47,* 168–181. https://doi.org/10.1016/j.chb.2014.09.034

Zijlstra-Shaw, S., & Stokes, C. W. (2018). Learning analytics and dental education; Choices and challenges. *European Journal of Dental Education, 22*(3), e658–e660. https://doi.org/10.1111/eje.12370

Zimmerman, B. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17. https://doi.org/10.1207/s15326985ep2501_2

Dimitrios Tzimas received the B.Sc degree in Informatics, Aristotle University of Thessaloniki, Greece, and the M.Sc degree in "Digital Signal Processing for Communications and Multimedia" from University of Athens, Greece. He is currently a Ph.D student at the Aristotle University of Thessaloniki. His research

interests include Learning Analytics and MOOCs. He has published so far six scientific articles in Greek conference proceedings. He teaches Computer Science in secondary and higher education for the last 20 years. He is a co-author of 4 Greek books about computer programming.

Stavros Demetriadis is currently Full Professor at the School of Informatics, Aristotle University of Thessaloniki (AUTh), Greece. He holds a B.Sc degree in Physics, M.Sc in Electronic Physics, and Ph.D in Multimedia educational technology from the Aristotle University of Thessaloniki. He teaches courses and researches in the broader area of Learning Technologies with emphasis on Computer-supported collaborative learning (CSCL), Conversational agents for learning, Adaptive Educational Hypermedia Systems, Educational robotics and Tangible interfaces for programming, Multimedia learning, Cognitive training technologies. He has published more than 150 research papers in international scientific journals (with IF) and international/national conference proceedings with more than 2000 third party citations (as of May 2018). He is a member of the scientific committee in several top-ranking international Conferences each year (such as IEEE ICALT, ECTEL, CSEDU, etc.). His supervised PhD project "Cubes Coding" was presented with awards in two international competitions (Open Education Challenge 2014 and NUMA-2014). His conference articles have received three times "Best paper" awards and one article has been highlighted as 'spotlight paper' at the IEEE Transactions of Learning Technologies journal. Since 1995 he has participated in many AUTH research projects and has been a project leader in some of them, such as the "Kaleidoscope Network of Excellence" AUTH team and the "colMOOC" international project to promote the use of conversational agents in MOOCs. He is also a Python enthusiast, offering the MOOC "Introduction to Programming with Python" (in Greek) and developing the 'pytolearn' website for advancing Python-based domain-related coding.