Routledge International Handbook of Schools and Schooling in Asia

Kerry J. Kennedy, John Chi-Kin Lee

Learning analytics

Publication details

PLEASE SCROLL DOWN FOR DOCUMENT

38

LEARNING ANALYTICS

Approaches and cases from Asia

*Bodong Chen, Chih-Ming Chen, Huang-Yao Hong, and
Ching Sing Chai*

Introduction

Learning analytics is an emerging and highly interdisciplinary field where many disciplines, such as education, computer science, and engineering, intersect. Since its first significant scholarly gathering in 2011, learning analytics has been mentioned increasingly in the news, technical reports, academic publications, and grant solicitations. The rise of this nascent field rests on a promise – and also a *premise* – that digital traces of learning can be turned into actionable knowledge to promote learning and teaching. As a young field, learning analytics is rooted in research areas such as business intelligence, user modeling, intelligent tutor systems (ITS), and social network analysis (Siemens, 2013). For instance, ITS, which have a strong presence in the learning analytics community, launched their first academic conference as early as 1998, while the analysis of social networks can be traced back to the dawn of the 20th century (Scott, 2013). That said, what is unique about learning analytics is its success in catalyzing broad discussions among various educational stakeholders (e.g., researchers, practitioners, administrators, policy-makers, and learners), leading to significant traction in academia, industry, and practice within a matter of years. Unlike many of its constituent areas, learning analytics has continued to foster the cross-fertilization of disciplines, which is often challenging but also extremely beneficial. More importantly, its strong intention to influence educational *practice* at multiple levels has nurtured awareness among practitioners, and thus the widespread publicity and reception of this field. As highlighted in the Horizon reports, learning analytics is a key trend in learning and teaching and has been widely adopted in recent years (Johnson et al., 2013; Johnson, Smith, Willis, Levine, & Haywood, 2011).

In this chapter, we present an overview of the field by articulating definitions and existing models of learning analytics. Case examples of learning analytics from Asian researchers are then summarized and reported. This is followed by an exploration of the key tensions in this field. The chapter concludes with a discussion of potential areas for future research in this area.

What is learning analytics?

The definition of learning analytics is plural and multifaceted because of the kaleidoscopic conceptions of *learning* and *analytics* introduced by various disciplines. The most widely adopted definition of learning analytics originated from its first conference: "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Long & Siemens, 2011, p. 34). This definition captures key components of learning analytics which are further articulated in a conceptual model comprising eight components, namely *collection*, *storage*, *cleaning*, *integration*, *analysis*, *representation*, *visualization*, and *action* (Siemens, 2013).

One approach to understanding learning analytics is through the lens of what learning analytics is *not*. Efforts have been made to distinguish learning analytics from its adjacent areas. *Academic analytics*, an area that was mostly inspired by business intelligence (Goldstein & Katz, 2005), is mentioned frequently together with learning analytics. In contrast to academic analytics, which is focused on institutions and administrators without much attention on pedagogy, learning analytics is more directly concerned with teachers and learners by attending to micro-patterns of learning (Long & Siemens, 2011). *Educational Data Mining* (EDM), a field that emerged a few years earlier, is also associated with learning analytics. According to Siemens (2013), while learning analytics is concerned with sense-making and action, EDM is more focused on developing methods specifically for the exploration of data that originate from educational settings. Although the techniques used are similar in the two fields, EDM has a more defined focus on reductionist analysis (Siemens & Baker, 2012), while learning analytics attends more to practice (Chatti, Dyckhoff, Schroeder, & Thüs, 2012). Despite these distinctions, however, it is expected that learning analytics, EDM, and academic analytics will continue to intersect and overlap in the future.

With *learning* and *analytics* as two pivotal concepts of the field, conceptions of learning analytics may stress one over the other. However, when a data- or analytic-centric perspective is taken, the phrase "learning analytics" often evokes images of data – especially "big data" – in education. For example, the *Horizon Report 2013* defines learning analytics as "education's approach to 'big data,' a science that was originally leveraged by businesses to analyze commercial activities, identify spending trends, and predict consumer behavior" (Johnson et al., 2013, p. 20). This emphasis on data is not surprising given that the abundance of data has been recognized as a key driver of this field (Long & Siemens, 2011). Such a data-centric view can also be observed in conceptual models of learning analytics. For instance, a model proposed by Chatti et al. (2012) includes *data collection and pre-processing*, *analytics and action*, and *post-processing* as three major components of learning analytics – without mentioning pedagogical contexts, sense-making, or intervention designs at all. This data-centric view of learning analytics is also seemingly magnified by the connection and co-evolution between learning analytics and massive open online courses (MOOCs), since MOOCs have produced "big" learning-related data (Kay, Reimann, Diebold, & Kummerfeld, 2013) and have attracted attention from learning analytics and EDM researchers (e.g., Coffrin, Corrin, Barba, & Kennedy, 2014).

However, scholars argue that conceptions of learning analytics need to recognize the nuanced aspects of *learning*, as learning analytics is essentially about learning (Gašević, Dawson, & Siemens, 2015). Suthers and Verbert (2013, p. 2) stress that "research on learning analytics may vary in the degree to which it makes technical contributions, but the connection to learning should be present." This point is often easily neglected in practical efforts to create learning analytics. Important arguments have been made that we need to contextualize specific analytics in their epistemological stances, pedagogy, and theories of assessment (Knight, Buckingham

Shum, & Littleton, 2014). Rather than treating analytics as being "pedagogically neutral" (e.g., Greller & Drachsler, 2012), an alternative approach is to recognize the highly nuanced nature of learning analytics and to coordinate the advances in both learning and analytics. Thus, despite the prominent growth of big data in recent years (Hey, Tansley, & Tolle, 2009), premature simplification of learning analytics – and overemphasis on data especially – can potentially do much harm to education in the absence of sufficient thinking about learning and education (Dringus, 2011). In order for learning analytics to play a constructive role in education, considerable work needs to be done on the learning side by designing pedagogically sound implementations (Wise & Vytasek, 2017).

Furthermore, learning analytics deals with educational phenomena at multiple levels. Work in the learning sciences has recognized that learning happens at multiple levels – not only in individual minds but also in small groups and larger learner communities (Stahl, 2013). For instance, collaborative knowledge-building as a group phenomenon depends on contributions from individuals but cannot be inferred reliably from individual learning (Scardamalia, Bransford, Kozma, & Quellmalz, 2012). This recognition of learning at multiple levels urges learning analytics to account for analytic levels beyond the individual level, which has been favored in traditional classroom settings, as well as in education regimes where accountability is the major concern. Expanding the focus from learning to a broader scope of education, Buckingham Shum (2012) differentiated the micro- (individual user or cohort), meso- (institution-wide), and macro- (region/state/national/international) levels of learning analytics. To assist decision-making at different levels, learning analytics also needs to attend to higher-level educational data, as well as to the integration and mutual complementarity between different levels.

To summarize, the meaning of learning analytics as a term is plural and multifaceted. Work in this field may emphasize *learning* and/or *analytics* in many different ways and at different levels. Such diverse understandings and approaches make it difficult to create a unified definition. Therefore, it is necessary for researchers in this field to provide an operational definition of the learning analytics they employ and how they relate the two key components.

Approaches and case examples of learning analytics studies in Asia

A review of learning analytics conducted by Asian researchers identified three commonly employed learning analytics approaches, namely (1) statistical sequential analysis, (2) social network analysis, and (3) data mining. The review was intended to be illustrative instead of comprehensive. These three approaches are defined briefly here. First, statistical sequential analysis – lag sequential analysis (LSA) in particular – is often used to examine whether or not a certain type of behavior increases or decreases the probability of an associated behavior. Social network analysis (SNA) can be used to analyze how individuals within a networked environment are connected through the various forms of online interactions and enables researchers to understand how network or group structures influence network/group performance. Lastly, data mining (DM) is employed to help make sense of the massive amounts of data available in and related to an education context. Cases of how these three approaches are utilized by Asian researchers in order to make sense of various forms of learning are presented next.

Lag sequential analysis

Lag sequential analysis (LSA), developed by Sackett (1979), is a statistical technique for analyzing sequential data. LSA enables one to explore, summarize, and statistically test cross-dependencies

between behaviors or events that occur in complex, interactive sequences (Faraone & Dorf-man, 1987). The word *lag* simply refers to LSA's interest in events or behaviors that follow each other; for example, lag=1 denotes an immediate sequence, and lag=2 refers to an inter-mediate temporal relationship with another event/behavior in the middle. The computational procedures of LSA generally proceed according to the following five steps: (1) calculate the sequential frequency transfer matrix, (2) calculate the condition probability matrix, (3) calculate the expected-value matrix, (4) calculate the adjusted residuals table, and (5) draw a sequential transfer diagram. By employing LSA, researchers can examine, for example, whether the presence of one discussion move (e.g., asking a question) increases the probability of another move (e.g., proposing an explanation) (Chen, Resendes, Chai, & Hong, 2017). Thus far, LSA has been applied to a wide range of learning contexts, such as information usage, online discussion, problem-solving, reading comprehension, gaming, and mobile learning. We provide a representative case in order to demonstrate how LSA research is conducted, followed by a summary of a variety of case examples.

The study by Lin, Hou, Wang, and Chang (2013) is selected as a representative case to explain how LSA is used to analyze sequential behavior patterns. The study employed LSA to examine the behavioral patterns of students' discussions in online Facebook forums, based on the quantitative content coding scheme of the Revised Bloom's Taxonomy (RBT). The RBT involves a two-dimensional structure involving a knowledge dimension and cognitive process dimension. The knowledge dimension includes four types of knowledge in the RBT, namely factual, conceptual, procedural, and metacognitive knowledge. The cognitive process dimen-sion contains six levels of cognitive process, namely remember, understand, apply, analyze, evaluate, and create. The unit of coding for performing LSA is a single message that students post or reply to in the Facebook group. The study applied the LSA to infer and visualize the sequence correlation of online discussion behaviors, thus depicting a clearer picture of the students' discussion behaviors on Facebook. The behavioral sequences refer to the order of the appearance of different behaviors. For instance, an online message coded as any of the six cognitive process dimensions may be followed by another message which could also be coded as any of the six categories. The LSA can help researchers to determine whether there are sig-nificant patterns in how the adjacent pairs of online messages are related. In other words, this method examines whether the appearance of one specific behavior followed by another specific behavior can reach statistical significance. By way of example, Lin et al.'s study reported that a message coded for cognitive process as "understand" is likely to be followed by another message that is in the same category based on the adjusted residual (z scores) computed through LSA.

In another similar use of LSA to analyze online collaborative discussions, Shukor, Tasir, Mei-jden, and Harun (2014) performed quantitative content analysis of students' learning behaviors (e.g., asking questions or sharing information) and levels of knowledge construction (e.g., high or low levels). The sequential transfer diagrams obtained through LSA clearly illustrate the dif-ferences between the high-level and low-level knowledge construction groups. The findings indicate that groups which were more effective in constructing high-level knowledge were more likely to negotiate shared information, while the presence of argumentation is also found to be useful for more successful knowledge construction.

Other than the application of LSA for online discussion, it has also been applied to unveil behavioral differences among learners with different learning preferences (Rezler & Rezmovic, 1981). Typically, such studies use questionnaires to distinguish students' learning preferences and to examine students' online behavioral differences through LSA. For example, Chen and Lin (2014) found that learners with a global learning preference out-performed those with a

sequential learning preference in the learning task assigned to them. In another study by Graf, Liu, and Kinshuk (2010), LSA was performed on learners' navigational behaviors in an online course. The authors found that students with different learning preferences tended to use different strategies to navigate online courses for more effective learning. These studies reveal that previously established psychological characterization of learners can be tested with LSA to determine if the theories hold true in the context of online behaviors.

Eye tracking as a means of collecting students' behavior data when they are interacting with computers has also been analyzed using LSA. For example, Tsai, Hou, Lai, Liu, and Yang (2012) used an eye-tracking technique to examine students' visual attention when they were performing science problem-solving. In particular, LSA was used to compare the scan patterns between successful and unsuccessful problem-solvers. A key finding showed that more successful (as compared with less successful) problem-solvers were better able to recognize and focus on relevant factors. In another study by Cao and Nishihara (2013), students' viewing behaviors when watching a video for deeper understanding of a given subject were captured with eye-tracking and analyzed using LSA. The main result showed that learners tended to give priority to the text rather than to the pictures.

LSA has also been used in contexts such as game-playing behaviors (e.g., Sung, Hou, Liu, & Chang, 2010) and mobile learning behaviors (e.g., Chiang, Yang, & Hwang, 2014). These studies clearly show the usefulness of LSA. Future possible applications of LSA in a variety of computer-based learning activities are, in a sense, limited only by the researchers' ability to code learning behaviors that are meaningful for study. Once the learning behaviors are coded, LSA can be performed to look for patterns in how behaviors are associated, and theories associated with such patterns can be built or refined. For example, writing with computers involves multiple steps which could be linked to students' writing strategies, misconceptions, and so on. How students act to complete a writing task can be analyzed by using LSA, and associated with different ways to profile students as writers. Similarly, students' creation of multimodal or computational materials could also be another area for LSA studies.

Social network analysis

Social network analysis (SNA) is an analytical method that looks into relationships among social entities and generates indices or metrics (e.g., the density and cohesion of a network, the centrality of an individual), as well as visualizations to describe them (Wasserman & Faust, 1994). When applied in education, SNA indices have been shown to be related to various aspects of learner experience (e.g., online learning satisfaction) and learning outcomes (e.g., deeper understanding of a topic inquired about online). Interest in applying SNA in education is motivated by an increasing focus on the "social" aspect of learning, demonstrated by learning theories emphasizing learning in groups (Stahl, 2006) and communities (Brown & Campione, 1994; Lave & Wenger, 1991; Scardamalia & Bereiter, 2003). We demonstrate the applications of SNA among Asian researchers in several research topics here.

First, SNA indices were employed as variables to measure how some forms of collaborative/cooperative learning affect the group processes. The study by Lin, Huang, and Chuang (2015) is selected to explain how SNA is used to analyze learners' interaction data generated in an e-learning environment. Graphs are a commonly used method of presenting interaction relationships in social networks. In the graph of a social network, a node represents a learner and a link represents a tie that exists between the pair of learners who interact. Generally, social networks can also be represented in the form of matrices. The simplest and most common matrix

is binary. In other words, if a tie is present, a "1" is entered into a cell of the matrix; if there is no tie, a "0" is entered. This kind of matrix is the starting point for almost all social network analysis and is referred to as an "adjacency matrix" because it represents who is next to, or adjacent to, whom, in the "social space," as mapped by the relations that have been measured. Another approach to measuring message flows is by distinguishing between "out-degree" and "in-degree" in a social network. Here, the out-degree refers to a situation in which an individual learner "A" asks another learner for assistance, whereas the in-degree implies that other learners ask "A" for assistance. According to Lin et al. (2015), a friendship network within a community is represented by using an undirected graph – that is, a friendship network does not consider in-/out-degree message flow. In a social network, network centrality can be measured further by three indices, namely degree centrality, closeness centrality, and betweenness centrality, based on the "adjacency matrix." A node with high centrality is generally in a more central position in a social network. Lin et al.'s study identified an individual's network position using degree centrality and examined how the level of self-regulation and the level of centrality simultaneously affect learning achievement in an SNA environment. Furthermore, whether both variables interacted with each other for learning achievement was also verified. Lin et al.'s study shows how network centrality affected online group learning behaviors, finding that groups that demonstrated a higher degree of network centrality would perform better in terms of learning achievements.

Other than learning achievements, SNA studies have generally shown that the social network indices are connected to the socio-emotional aspect of learning. Lee and Bonk (2016) employed SNA to examine learner relationships and their interactions in an online course using weblogs that were designed for writing and sharing reflective journals. The results showed that learners' blogging activities (as measured mainly by the online interactions among the learners, e.g., the change of in-degree and out-degree in online interactions) had significant pre-post learning impacts (including learners' perceived emotional closeness, a positive correlation between online interactions and peer closeness network values, and change of social network structure) during a 16-week graduate course. In another longitudinal study spanning 2006 to 2012 (Lin, Hu, Hu, & Liu, 2016), which examined the relationship between face-to-face and online collaborations by a group of K-12 teachers (N=172), several SNA techniques (such as sociograms, centrality, cohesive subgroups, and the clique phenomenon) were applied. The study revealed that face-to-face and online collaborations are both essential in teaching and continuously complement one another in terms of teachers' professional development.

SNA has been used widely as a pedagogical tool to foster group-based learning. Chen and Chang's (2014) study is an example of using social interaction information in web-based collaborative learning environments to guide learners in identifying suitable learning partners for more effective collaboration. Hong, Scardamalia, Messina, and Teo (2015) explored students' autonomous use of three analytic tools – a Vocabulary Analyzer, a Social Network Tool, and a Semantic Overlap Tool – to improve community knowledge-building in a class of Grade 5/6 students. It was found that these knowledge-building analytic tools are useful in helping the students to form a more discursively connected community. As a consequence, students become progressively more engaged in deeper inquiry.

As the social dimensions of learning and its associated learning theories gain attention among researchers, and as the use of social media continues to develop, SNA is likely to become increasingly important for online platforms. Future research on how teachers and online community members consciously design or apply social strategies that foster the social dimension of learning can be conducted by using the SNA indices as dependent variables.

LSA can be the means to track the patterns of social behaviors that denote a member's social moves.

Data mining

Data mining is defined as the process of discovering meaningful patterns in data (Witten & Frank, 2005, p. 5). Data mining techniques such as decision trees, a support tool that employs a tree-like graph of decisions with possible consequences, have been investigated in association with some established variables in education, such as cognitive styles, prior knowledge, reasoning processes, and learning outcomes (Doleck, Basnet, Poitras, & Lajoie, 2015; Romero, Espejo, Zafra, Romero, & Ventura, 2013). We explicate a representative case here, followed by a summary of relevant studies using data mining.

Chen and Huang (2013) utilized data mining techniques (e.g., K-means algorithm) of clustering to investigate how prior knowledge affects game-based learning. They found that prior knowledge has positive impacts on the type of games which foster procedural knowledge rather than declarative knowledge. Their study explains how data mining techniques are applied to conduct data analyses in educational settings in order to discover hidden relationships that may be ignored using more traditional statistical methods. Frequently used data mining techniques include classification, clustering, association, and prediction. Among various data mining techniques, Chen and Huang adopted the K-means algorithm – a widely used clustering technique to group learners into several sub-groups with similar characteristics. They selected students' pre-test scores and post-test scores and their responses to the questionnaire (which indicate the frequencies of playing digital games and prior experience with digital games) as the grouping attributes, and the K-means algorithm was applied to group the students. The differences between the five clusters generated were then described using the mean scores and standard deviation for each of the attributes. Additionally, frequency counts and percentages were applied to explain the distribution of the participants among the clusters. After obtaining the five clusters, analysis of variance (ANOVA) was used to identify whether significant differences existed among the five clusters in terms of their learning performances and their learning perceptions. The outcome shows that there are significant differences among the five clusters in the two aspects tested.

Similarly, Chen and Liu (2008) utilized data mining techniques (a decision tree based on eight formulated rules) along with some traditional statistics to explore whether cognitive styles (i.e., learners' constant latent tendency of information processing modes, as defined by Messick, 1984) have any impact on students' learning patterns in a web-based instruction (WBI) program. The findings confirmed this to be the case.

As with SNA, data mining techniques have also been employed to improve actual teaching and learning. For example, Chen, Hsieh, and Hsu (2007) mined learners' profiles to detect their common misconceptions and then proposed and tested a remediation approach to enhance students' learning performance based on the learners' profiles. In another experimental study by Lin, Yeh, Hung, and Chang (2013), a personalized creativity learning system was designed and developed using decision trees, with the purpose of optimizing learners' creativity performance by providing them with personalized learning paths. The findings based on the experiment indicated that the learners have a much higher probability of obtaining a good creativity score when a learning path suggested by a decision tree is available to them. However, the application of data mining techniques to optimize learning raises questions of learner agency, and we explore this in a later section.

Data mining represents an emerging and innovative means of discovering learning patterns in large educational data sets and, as such, offers the potential for a number of new research areas. One possible research direction is to capitalize on the strengths of various data mining approaches to further advance the development of online learning (e.g., MOOCs). Another is concerned with integrating data mining into institutional research. As an increasing number of schools, colleges, and universities are concerned with making informed decisions about their management on a large scale (e.g., admissions, curriculum planning, and student life), data mining is likely to become more important for the educational industry in the future as well.

Key tensions in the field of learning analytics

While the previously mentioned research provides substantial demonstration of the usefulness of analytical approaches applied in learning analytics, the interdisciplinary nature of learning analytics has given rise to various tensions of which further development in this field needs to be aware. Attending to those tensions is necessary for effective and ethical design, implementation, and use of learning analytics. Salient tensions we have observed from the literature to date include the tensions (1) among various conceptions of learning, (2) between learning and

Table 38.1 Summary of some emerging research themes/topics in the learning analytics area with selected study examples

Theme	Some key topics studied or variables measured/assessed	Selected related studies
Lag sequential analysis (LSA): an analytic technique for processing sequential event data	Learners' discussion behaviors in online (discussion) forums	– Lin et al. (2013) – Shukor et al. (2014) – Chen and Lin (2014)
	Behavioral differences among learners with different learning styles	– Graf et al. (2010)
	Eye-tracking research	– Tsai et al. (2012) – Cao and Nishihara (2013)
	Other topics, such as inquiry learning, game-playing, and mobile learning behaviors	– Hou (2015) for inquiry behaviors – Sung et al. (2010) for game-playing behaviors – Chiang et al. (2014) for mobile learning behaviors
Social network analysis (SNA): an analysis for constructing, measuring, or visualizing networks based on relations among network "members"	Group learning behaviors	– Lin et al. (2015) – Lin et al. (2016) – Lee and Bonk (2016)
	To promote collaborative learning	– Chen and Chang (2014)
	To promote community knowledge advancement	– Hong et al. (2015)
Data mining (DM): a general analytic technique to extract or discover patterns of certain variables in "big" data sets	To identify cognitive styles for web-based learning	– Chen and Liu (2008) – Chen and Huang (2013)
	To improve teaching and learning	– Chen et al. (2007) – Lin et al. (2013)

computer algorithms, (3) between agency and control, and (4) surrounding ethical access and use of educational data. These are elaborated upon here.

Different conceptions of learning

Learning analytics is not only about a set of analytic techniques but is also concerned with feeding results to relevant stakeholders in order to improve educational practice. On the surface, what we consider as learning would dictate the types of data collected for the analysis of "learning." Furthermore, researchers may hold different epistemological assumptions – about where knowledge resides and how knowledge develops – and will thus consider learning in significantly different ways. For instance, a researcher who conceptualizes learning as knowledge acquisition in individual minds is more likely to emphasize test scores as hallmarks of academic performance (e.g., Agudo-Peregrina et al., 2014). By contrast, a researcher who views knowledge as dependent on social interactions would look into group activities for indicators of learning (Buckingham Shum & Deakin Crick, 2012). Recently, Koh et al. (2015) adopted the constructivist-oriented meaningful learning framework (see Howland, Jonassen, & Marra, 2012) and attempted to aggregate indices generated by SNA and students' participation data in order to provide feedback for teachers and students with regards to their meaningful engagement. These examples illustrate that while different conceptions of learning commonly exist in many education research communities, the divide is even wider when educational theorists, computer scientists, and engineers interact with one another. As we go forward, any effort in creating learning analytics would require that, at the very least, researchers and developers make their conceptions of learning explicit in order to ensure that all involved share a common point of departure.

Learning and computer algorithms

Learning analytics is sometimes conceptualized as translating digital traces into numbers for the purpose of interpretation. For example, it is compared to immersing a thermometer in water to gauge the temperature (Greller & Drachsler, 2012). Learning analytics, in this case, is considered to be "pedagogically neutral." However, this is hardly the case when making sense of learning from learners' "digital shadows" in learning environments (Buckingham Shum, 2015). During the process, critical decisions are made regarding data collection, choice of algorithms, interpretation of results, and possible courses of action from the results of learning analytics. These decisions are laden with values and beliefs about learning. From a different perspective, any results of computer algorithms are contingent upon informed and contextualized interpretations; real changes in education cannot be brought about by simply giving people volumes of numbers generated by learning analytics (Macfadyen & Dawson, 2012). Thus, oversimplification of what constitutes learning may lead to misinformed practices that neglect important aspects of learning.

Siemens (2013, p. 1395) cautions learning analysts to "keep human and social processes central in learning analytics activities," as the learning process is essentially social and creative. A similar caution is made by Gašević et al. (2015), who have problematized practices of counting certain activities for correlation with academic performance. Instead, these researchers call for work around coherent theoretical models of learning behaviors. It is especially problematic if algorithms exist in "black boxes" and produce numbers that are in the hands of those with the power to influence behaviors (Buckingham Shum, 2015) in the name of "optimizing" learning. This tension between learning and algorithms will continue to figure in learning analytics.

Agency and control

Learning analytics influences learners' agency when it is used to shape how learning is interpreted and how learners need to act (Buckingham Shum, 2015). There is a constant tension between analytics and learners because the use of analytics could take way agency from the learners. Analytics could disempower learners, making them increasingly reliant on institutions to provide them with continuous feedback, rather than developing meta-cognitive and learning-to-learn skills and dispositions (Buckingham Shum & Ferguson, 2012). This is especially the case when the analytics is contained in "black boxes" and when learners are given results to react to without understanding how the algorithms work. The popularity of dashboards in learning analytics rests on their power to translate complex data into digestible visualizations; however, their designs range from a traffic signal to complex interactive visualizations (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013), reflecting different levels of empowerment and control.

In addressing this tension, learner agency has been recognized as an important principle in designing learning analytics interventions (Wise, 2014). A special issue of the *Journal of Learning Analytics* features the application of learning analytics to promoting 21st-century competencies, which often emphasize greater agency of learners (Buckingham Shum & Deakin Crick, 2016). The balance between control and agency is a high-stakes endeavor, such that work in learning analytics will not undo decades of work in education to promote high-level agencies that are important for people in the knowledge age. In addition to facilitating administration, awareness, reflection, and sense-making, learning analytics should invest in assisting learners and stakeholders to develop local decision-making rather than taking such power away from them.

Data access and ethics

Ethical issues related to learning analytics are multi-dimensional. First, learning analytics researchers are divided on who owns the data and whether usage of certain learning platforms should incur consent of data use for analytical purposes. For instance, Macfadyen and Dawson (2012, p. 151) note that "data that is gathered through institutional research is subject to the provisions of the Freedom of Information and Protection of Privacy Act," rather than the traditional ethics review process for academic research. While it may be unreasonable to require researchers to obtain consent from all learners, research ethics need to be evaluated regularly, even if data are already made accessible (Boyd & Crawford, 2012).

Ethical issues may figure more deeply in power relations between educational institutions and stakeholders. In higher education, for instance, power relations among students and universities should be brought to bear when designing learning analytics (Slade & Prinsloo, 2013), as they may lead to important consequences (e.g., graduation) for individuals. In classrooms, the ethical use of learning analytics needs to consider the vulnerability of students to protect them from possible harm. For example, publicly comparing all students' performance may undermine the confidence of low-performing students, while presenting individual performance with a class average may demotivate middle- to high-performing students who could actually invest more effort in their learning. The impact of power relations stretches from data collection to how analytics could be interpreted and have real-life, powerful impacts on individuals.

Finally, it should be noted that the tensions outlined here are interconnected and thus should not be viewed as separate concerns. For example, the moral practice of learning analytics depends not only on the ethical use of data but also on moral applications that look beyond academic "effectiveness" (Slade & Prinsloo, 2013). Critically exploring different conceptions of

learning would lead to more sensible discussion of agency and control when designing learning analytics. Kaleidoscopic as the field is, those tensions will exist as long as there is participation from multiple disciplines. The goal is not to achieve a unified view of the field, but rather to create a meaningful common ground for further development.

Conclusions

Learning analytics is an emerging field of study that has much potential to contribute to the understanding and enrichment of learning. It draws on rich sources of data that are, for example, generated by the learners when they are performing learning tasks on computer-based platforms. Such data are richer, more fine-grained, and arguably more "objective" in comparison to students' self-reported data and even classroom observation data. As illustrated in this chapter, such fine-grained process data can be associated with many existing learning theories and, in turn, produce legitimate new evidence to fuel refinements of learning theories. However, we believe that there is no value-free data collection and interpretation process. As more techniques pertaining to learning analytics emerge, the discussion of wise, ethical, and valid uses of learning analytics has to develop concurrently.

In terms of future research on the use of learning analytics, we expect more cross-fertilization among analytical techniques when studying the same learning phenomenon. For instance, researchers could apply two or more of the approaches reviewed here concurrently in a given study, such as using SNA together with LSA in order to understand how social network formation is related to patterns of learning behaviors and activities. The intersection of these techniques may shed new light on understanding user behaviors. Another area that needs attention is users' (such as teachers and learners) responses to the analytics reports provided to them. Learning analytics should allow teachers to analyze students' learning in a useful manner and to promote students' self-directed and/or collaborative learning or other forms of generative learning behaviors. However, the end users need to possess the necessary literacies required to understand the feedback provided through the analytics in order for the feedback to be useful. There may be a need, therefore, to study the effect of levels of literacy users possess with regards to learning analytics, how this influences the users in making sense of the feedback, and, more important, how educators could help them to develop the required literacy in their educational contexts. Lastly, each tool added to a learning environment perturbs the ecology. Understanding how learning analytics creates new activity systems (see Engestrom, 2000) is likely to be an important area for future research. Substantial efforts are needed to resolve the aforementioned tensions in different activity systems.

References

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior*, *31*, 542–550. doi: 10.1016/j.chb.2013.05.031

Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society*, *15*(5), 662–679. doi: 10.1080/1369118X.2012.678878

Brown, A. L., & Campione, J. C. (1994). Guided discovery in a community of learners. In K. McGilly (Ed.), *Classroom lessons: Integrating cognitive theory and classroom practice*. Cambridge, MA: MIT Press/ Bradford Books.

Buckingham Shum, S. B. (2012). *UNESCO Policy Brief: Learning Analytics* (No. November). UNESCO Institute for Information Technologies in Education. Retrieved from www.iite.unesco. org/publications/3214711/

Buckingham Shum, S. B. (2015). Learning analytics: White rabbits and silver bullets. In *Coding/learning: Software and digital data in education* (pp. 44–51). Stirling: University of Stirling.

Buckingham Shum, S. B., & Deakin Crick, R. (2012). Learning dispositions and transferable competencies. In *Proceedings of the 2nd international conference on learning analytics and knowledge – LAK '12* (p. 92). New York, NY: ACM Press. doi: 10.1145/2330601.2330629

Buckingham Shum, S., & Deakin Crick, R. (2016). Learning Analytics for 21st century competencies. Journal of Learning Analytics, 3(2), 6–21. doi: 10.18608/jla.2016.32.2

Buckingham Shum, S., & Ferguson, R. (2012). Social learning analytics. *Educational Technology and Society*, *15*(3), 3–26.

Cao, J., & Nishihara, A. (2013). Viewing behaviors affected by slide features and learning style in slide video from a sequence analysis perspective. *The Journal of Information and Systems in Education*, *12*(1), 1–12.

Chatti, M. A., Dyckhoff, A. L., Schroeder, U., & Thüs, H. (2012). A reference model for learning analytics. *International Journal of Technology Enhanced Learning*, *4*(5), 318–331.

Chen, B., Resendes, M., Chai, C. S., & Hong, H.-Y. (2017). Two tales of time: Uncovering the significance of sequential patterns among contribution types in knowledge-building discourse. *Interactive Learning Environments*, *25*(2), 162–175. https://doi.org/10.1080/10494820.2016.1276081

Chen, C.-M., & Chang, C.-C. (2014). Mining learning social networks for cooperative learning with appropriate learning partners in a problem-based learning environment. *Interactive Learning Environments*, *22*(1), 97–124. doi: 10.1080/10494820.2011.641677

Chen, C.-M., Hsieh, Y.-L., & Hsu, S.-H. (2007). Mining learner profile utilizing association rule for web-based learning diagnosis. *Expert Systems with Applications*, *33*(1), 6–22. doi: 10.1016/j.eswa.2006.04.025

Chen, S. Y., & Huang, P.-R. (2013). The comparisons of the influences of prior knowledge on two game-based learning systems. *Computers & Education*, *68*, 177–186. doi: 10.1016/j.compedu.2013.05.005

Chen, C.-M., & Lin, S.-T. (2014). Assessing effects of information architecture of digital libraries on supporting E-learning: A case study on the digital library of nature & culture. *Computers & Education*, *75*, 92–102. doi: 10.1016/j.compedu.2014.02.006

Chen, S. Y., & Liu, X. (2008). An integrated approach for modeling learning patterns of students in Web-based instruction: A cognitive style perspective. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *15*(1), 1. doi:10.1145/1352782.1352783

Chiang, T. H. C., Yang, S. J. H., & Hwang, G.-J. (2014). Students' online interactive patterns in augmented reality-based inquiry activities. *Computers & Education*, *78*, 97–108. doi: 10.1016/j.compedu.2014.05.006

Coffrin, C., Corrin, L., Barba, P. de, & Kennedy, G. (2014). Visualizing patterns of student engagement and performance in MOOCs. In *Proceedings of the fourth international conference on learning analytics and knowledge – lAK '14* (pp. 83–92). New York, NY: ACM Press. doi: 10.1145/2567574.2567586

Doleck, T., Basnet, R. B., Poitras, E. G., & Lajoie, S. P. (2015). Mining learner – system interaction data: Implications for modeling learner behaviors and improving overlay models. *Journal of Computers in Education*, *2*(4), 421–447. doi: 10.1007/s40692-015-0040-3

Dringus, L. P. (2011). Learning analytics considered harmful. *Journal of Asynchronous Learning Networks*, *16*(3), 87–101.

Engestrom, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics*, *43*(7), 960–974.

Faraone, S. V., & Dorfman, D. D. (1987). Lag sequential analysis: Robust statistical methods. *Psychological Bulletin*, *101*(2), 312–323. https://doi.org/10.1037//0033-2909.101.2.312

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, *59*(1), 64–71. https://doi.org/10.1007/s11528-014-0822-x

Goldstein, P. J., & Katz, R. N. (2005). *Academic analytics: The uses of management information and technology in higher education.* Boulder, CO: EDUCAUSE Center for Applied Research.

Graf, S., Liu, T.-C., & Kinshuk. (2010). Analysis of learners' navigational behaviour and their learning styles in an online course. *Journal of Computer Assisted Learning*, *26*(2), 116–131. doi: 10.1111/j.1365–2729.2009.00336.x

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, *15*(3), 42–57.

Hey, A. J. G., Tansley, S., & Tolle, K. M. (2009). *The fourth paradigm: Data-intensive scientific discovery.* Microsoft Research Redmond, WA.

Hong, H.-Y., Scardamalia, M., Messina, R., & Teo, C. L. (2015). Fostering sustained idea improvement with principle-based knowledge building analytic tools. *Computers & Education*, *89*, 91–102.

Hou, H.-T. (2015). Integrating cluster and sequential analysis to explore learners' flow and behavioral patterns in a simulation game with situated-learning context for science courses: A video-based process exploration. *Computers in Human Behavior*, *48*, 424–435. doi: 10.1016/j.chb.2015.02.010

Howland, J. L., Jonassen, D. H., & Marra, R. M. (2012). *Meaningful learning with technology*. Upper Saddle River, NJ: Pearson.

Jiang, M., Cui, P., & Faloutsos, C. (2016). Suspicious behavior detection: Current trends and future directions. *Intelligent Systems, IEEE*, *31*(1), 31–39.

Johnson, L., Adams Becker, S., Cummins, M. V. E., Freeman, A., & Ludgate, H. (2013). *NMC horizon report: 2013 K-12 edition*. Austin, TX: The New Media Consortium. Retrieved from www.nmc.org/pdf/2013-horizon-report-k12.pdf

Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). *The 2011 horizon report*. Austin, TX: The New Media Consortium.

Kay, J., Reimann, P., Diebold, E., & Kummerfeld, B. (2013). MOOCs: So many learners, so much potential . . . *IEEE Intelligent Systems*, *28*(3), 70–77. doi: 10.1109/MIS.2013.66

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics*, *1*(2), 23–47.

Koh, A. L. H., Wong, K. M., & Chai, C. S. (2015). Tracking quality of learning through analytic for meaningful learning with online platform: An initial conceptualization. In A.-F. Lai, C. S. Chai, X. Gu, Y.-T. Wu, & L.-J. Zhang (Eds.), *Proceedings of the global Chinese conference on computers in education: Teacher forum* (pp. 102–105). Taipei, Taiwan: GCSCE.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lee, J., & Bonk, C. J. (2016). Social network analysis of peer relationships and online interactions in a blended class using blogs. *The Internet and Higher Education*, *28*, 35–44. doi: 10.1016/j.iheduc.2015.09.001

Lin, C. F., Yeh, Y., Hung, Y. H., & Chang, R. I. (2013). Data mining for providing a personalized learning path in creativity: An application of decision trees. *Computers & Education*, *68*, 199–210. doi: 10.1016/j.compedu.2013.05.009

Lin, J.-W., Huang, H.-H., & Chuang, Y.-S. (2015). The impacts of network centrality and self-regulation on an e-learning environment with the support of social network awareness. *British Journal of Educational Technology*, *46*(1), 32–44. doi: 10.1111/bjet.12120

Lin, P.-C., Hou, H.-T., Wang, S.-M., & Chang, K.-E. (2013). Analyzing knowledge dimensions and cognitive process of a project-based online discussion instructional activity using Facebook in an adult and continuing education course. *Computers & Education*, *60*(1), 110–121. doi: 10.1016/j.compedu.2012.07.017

Lin, X., Hu, X., Hu, Q., & Liu, Z. (2016). A social network analysis of teaching and research collaboration in a teachers' virtual learning community. *British Journal of Educational Technology*, *47*(2), 302–319.

Long, P., & Siemens, G. (2011). Penetrating the Fog: Analytics in learning and education. *Educause Review*, *46*(5), 30–32.

Macfadyen, L., & Dawson, S. (2012). Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Journal of Educational Technology & Society*, *15*(3), 149–163.

Messick, S. (1984). The nature of cognitive styles: Problems and promise in educational practice. *Educational Psychologist*, *19*(1), 59–74.

Rezler, A. G., & Rezmovic, V. (1981). The learning preference inventory. *Journal of Allied Health*, *10*(1), 28–34.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135–146. doi: 10.1002/cae.20456

Sackett, G. P. (1979). The lag sequential analysis of contingency and cyclicity in behavioral interaction research. In J. D. Osofsky (Ed.), *Handbook of infant development*. New York: Wiley.

Scardamalia, M., & Bereiter, C. (2003). Knowledge building. In J. W. Guthrie (Ed.), *Encyclopedia of education* (2nd ed., Vol. 17, pp. 1370–1373). New York, NY: Macmillan Reference.

Scardamalia, M., Bransford, J. D., Kozma, B., & Quellmalz, E. (2012). New assessments and environments for knowledge building. In *Assessment and teaching of 21st century skills* (pp. 231–300). The Netherlands: Springer. doi:10.1007/978-94-007-2324-5_5

Scott, J. (2013). *Social network analysis* (3rd ed.). London, UK: Sage.

Shukor, N. A., Tasir, Z., Meijden, H. V. der, & Harun, J. (2014). Exploring students' knowledge construction strategies in computer-supported collaborative learning discussions using sequential analysis. *Journal of Educational Technology & Society*, *17*(4), 216–228.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist*, *57*(10), 1380–1400.

Siemens, G., & Baker, R. S. J. d. (2012). Learning analytics and educational data mining: Towards communication and collaboration. In *Proceedings of the 2nd international conference on learning analytics and knowledge* (pp. 252–254). New York: ACM.

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

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Stahl, G. (2013). Learning across levels. *International Journal of Computer-Supported Collaborative Learning*, *8*(1), 1–12.

Sung, Y.-T., Hou, H.-T., Liu, C.-K., & Chang, K.-E. (2010). Mobile guide system using problem-solving strategy for museum learning: A sequential learning behavioural pattern analysis. *Journal of Computer Assisted Learning*, *26*(2), 106–115. doi: 10.1111/j.1365-2729.2010.00345.x

Suthers, D., & Verbert, K. (2013). Learning analytics as a "middle space". In *Proceedings of the third international conference on learning analytics and knowledge – lAK '13* (pp. 1–4). New York, NY: ACM Press. doi: 10.1145/2460296.2460298

Tsai, M.-J., Hou, H.-T., Lai, M.-L., Liu, W.-Y., & Yang, F.-Y. (2012). Visual attention for solving multiple-choice science problem: An eye-tracking analysis. *Computers & Education*, *58*(1), 375–385. doi: 10.1016/j.compedu.2011.07.012

Verbert, K., Duval, E., Klerkx, J., Govaerts, S., & Santos, J. L. (2013). Learning analytics dashboard applications. *American Behavioral Scientist*, *57*(10), 1500–1509. doi: 10.1177/0002764213479363

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge, UK: Cambridge University Press.

Wise, A. F. (2014). Designing pedagogical interventions to support student use of learning analytics. In *Proceedings of the fourth international conference on learning analytics and knowledge – LAK '14* (pp. 203–211). New York, NY: ACM Press. doi: 10.1145/2567574.2567588

Wise, A. F., & Vytasek, J. (2017). Learning analytics implementation design. In C. Lang, G. Siemens, A. F. Wise, & D. Gašević (Eds.), *Handbook of learning analytics* (pp. 151–160). Society for Learning Analytics Research.

Witten, I. H., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. San Francisco, CA: Morgan Kaufmann.