

Learning Analytics: Challenges and Future Research Directions

Mohamed Amine Chatti
chatti@informatik.rwth-aachen.de

Vlatko Lukarov
Hendrik Thüs
Arham Muslim
Ahmed Mohamed Fahmy Yousef
Usman Wahid
Christoph Greven
Arnab Chakrabarti
Ulrik Schroeder

RWTH Aachen University
Informatik 9 (Learning Technologies)

urn:nbn:de:0009-5-40350

Zusammenfassung

In den letzten Jahren hat Learning Analytics (LA) viel Aufmerksamkeit im Bereich der technology-enhanced learning (TEL) Forschung auf sich gezogen, da Anwender, Institutionen und Forscher zunehmend das Potenzial sehen, das LA hat um die Zukunft der TEL Landschaft zu gestalten. Generell beschäftigt LA sich mit der Entwicklung von Methoden, die Bildungsdatensätze nutzbar zu machen um den Lernprozess zu unterstützen. Dieses Manuskript bietet eine Grundlage für die zukünftige Forschung im Bereich LA. Es bietet einen systematischen Überblick über dieses aufstrebende Gebiet und seine Schlüsselkonzepte durch ein Referenzmodell für LA, welches auf vier Dimensionen basiert, namentlich Daten, Umgebungen und Kontext (Was?), Akteure (Wer?), Ziele (Warum?) und Methoden (Wie?). Darüber hinaus identifiziert es verschiedene Herausforderungen und Forschungsmöglichkeiten im Bereich LA in Bezug auf jede dieser Dimensionen.

Stichwörter: e-learning, learning analytics, educational data mining, personalization, seamless learning, context modeling, lifelong learner modeling, open assessment

Abstract

In recent years, learning analytics (LA) has attracted a great deal of attention in technology-enhanced learning (TEL) research as practitioners, institutions, and researchers are increasingly seeing the potential that LA has to shape the future TEL landscape. Generally, LA deals with the development of methods that harness educational data sets to support the learning process. This paper provides a foundation for future research in LA. It provides a systematic overview on this emerging field and its key concepts through a reference model for LA based on four dimensions, namely data, environments, context (what?), stakeholders (who?), objectives (why?), and methods (how?). It further identifies various challenges and research opportunities in the area of LA in relation to each dimension.

Keywords: e-learning, learning analytics, educational data mining, personalization, seamless learning, context modeling, lifelong learner modeling, open assessment

1. Introduction

Learning is increasingly distributed across space, time, and media. Consequently, a large volume of data about learners and learning is being generated. This data is mainly traces that learners leave as they interact with networked learning environments. In the last few years, there has been an increasing interest in the automatic analysis of educational data to enhance the learning experience, a research area referred to recently as learning analytics. The most commonly-cited definition of learning analytics, which was adopted by the first international conference on learning analytics and knowledge (LAK11), 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” (as cited in Siemens & Long, 2011, Learning Analytics section, para. 2).

Learning analytics is not a genuine new research area. It reflects a field at the intersection of numerous academic disciplines (e.g. learning science, pedagogy, psychology, Web science, computer science) (Dawson et al., 2014). It borrows from different related fields (e.g. academic analytics, action analytics, educational data mining, recommender systems, personalized adaptive learning) and synthesizes several existing techniques (e.g. machine learning/artificial intelligence, information retrieval, statistics and visualization) (Chatti et al., 2012a; Ferguson, 2012).

The emerging field of learning analytics is constantly developing new ways to analyze educational data. However, there is a limited application of analytics to inform practice within the education sector (Dawson et al., 2014; Manyika et al., 2011). Moreover, relatively little research has been conducted so far in order to provide a systematic overview of the field and to identify various challenges that lay ahead, as well as future research directions in this fast-growing area. Thus, there is a need to highlight the potential for learning analytics to improve the learning and teaching processes. In this paper, we propose a four-dimensional reference model for LA that can foster a common understanding of key concepts in this emerging field. We then build on this model to identify a series of challenges and develop a number of insights for learning analytics research in the future.

2. Learning Analytics

Different definitions have been provided for the term Learning Analytics (LA). Ferguson (2012) and Clow (2013a) compile some LA definitions and provide a good overview on the evolution of LA in recent years. Although different in some details, LA definitions share an emphasis on converting educational data into useful actions to foster learning.

Furthermore, it is noticeable that these definitions do not limit LA to automatically conducted data analysis. In this paper, we view LA as a technology-enhanced learning (TEL) research area that focuses on the development of methods for analyzing and detecting patterns within data collected from educational settings, and leverages those methods to support the learning experience.

LA concepts and methods are drawn from a variety of related research fields including academic analytics, action analytics, and educational data mining. In general, academic analytics and action analytics initiatives applied analytics methods to meet the needs of

educational institutions. They only focused on enrolment management and the prediction of student academic success and were restricted to statistical software (Goldstein and Katz, 2005; Campbell et al., 2007; Norris et al., 2008). Besides just serving the goals of educational institutions, applications of LA can be oriented toward different stakeholders including learners and teachers. Beyond enrolment management and prediction, LA is increasingly used to achieve objectives more closely aligned with the learning process, such as reflection, adaptation, personalization, and recommendation. In addition to just providing statistics, more recent LA approaches apply several other analytics methods such as data mining to guide learning improvement and performance support.

The analysis domain, data, process and objectives in LA and Educational Data Mining (EDM) are quite similar (Siemens & Baker, 2012). Both fields focus on the educational domain, work with data originating from educational environments, and convert this data into relevant information with the aim of improving the learning process. However, the focus in LA can be different from the one in EDM. EDM has a more technological focus, as pointed out by Ferguson (2012). It basically aims at the development and application of data mining techniques in the educational domain (data-driven analytics). LA, on the other hand, has a more pedagogical focus. It puts different analytics methods into practice for studying their actual effectiveness on the improvement of teaching and learning (learner-focused analytics). As Clow (2013a) puts it: “Learning analytics is first and foremost concerned with learning” (p. 687).

3. A Reference Model for Learning Analytics

In this section, we propose a reference model for LA based on four dimensions. Our endeavor is to give a systematic overview of LA and its related concepts that will support communication between researchers as they seek to address the various challenges that will arise as understanding of the technical and pedagogical issues surrounding learning analytics evolves.

As depicted in Figure 1, the four dimensions of the proposed reference model for LA are:

- **What?** What kind of data does the system gather, manage, and use for the analysis?
- **Who?** Who is targeted by the analysis?
- **Why?** Why does the system analyze the collected data?
- **How?** How does the system perform the analysis of the collected data?

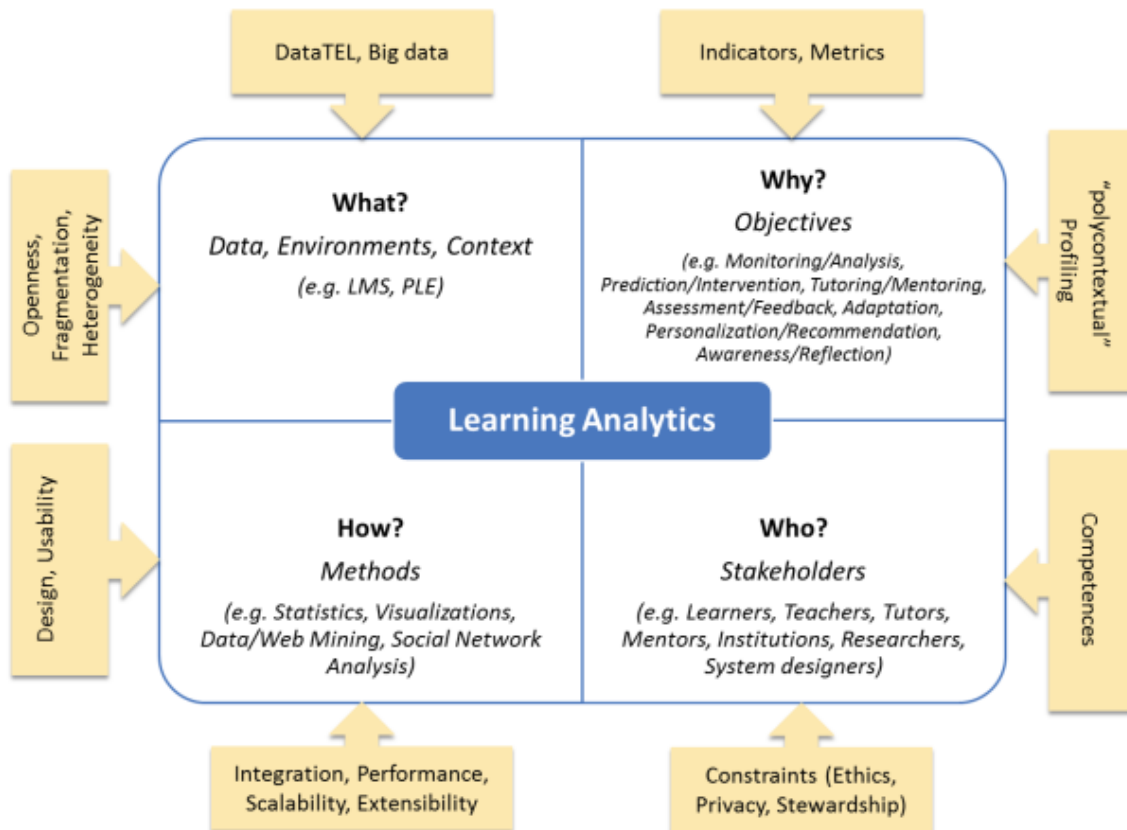


Figure 1. Learning Analytics Reference Model

3.1. Data, Environments, Context (What?)

LA is a data-driven approach. The “what?” dimension refers to the kind of data used in the LA task. It also refers to the environments and contexts in which learning occurs. An interesting question in LA is where the educational data comes from. LA approaches use varied sources of educational data. These sources fall into two big categories: centralized educational systems and distributed learning environments. Centralized educational systems are well represented by learning management systems (LMS). LMS accumulate large logs of data of the students’ activities and interaction data, such as reading, writing, accessing and uploading learning material, taking tests, and sometimes have simple, built-in reporting tools (Romero and Ventura, 2007). LMS are often used in formal learning settings to enhance traditional face-to-face teaching methods or to support distance learning.

User-generated content, facilitated by ubiquitous technologies of production and cheap tools of creativity, has led to a vast amount of data produced by learners across several learning environments and systems. With the growth of user-generated content, LA based on data from distributed sources is becoming increasingly important and popular. Open and distributed learning environments are well represented by the personal learning

environment (PLE) concept (Chatti et al., 2012b). PLEs compile data from a wide variety of sources beyond the LMS. The data comes from formal as well as informal learning channels. It can also come in different formats, distributed across space, time, and media.

3.2. Stakeholders (Who?)

The application of LA can be oriented toward different stakeholders, including students, teachers, (intelligent) tutors/mentors, educational institutions (administrators and faculty decision makers), researchers, and system designers with different perspectives, goals, and expectations from the LA exercise. Students will probably be interested in how analytics might improve their grades or help them build their personal learning environments. Teachers might be interested in how analytics can augment the effectiveness of their teaching practices or support them in adapting their teaching offerings to the needs of students. Educational institutions can use analytics tools to support decision making, identify potential students “at risk”, improve student success (i.e. student retention and graduation rates) (Campbell et al., 2007), develop student recruitment policies, adjust course planning, determine hiring needs, or make financial decisions (Educause, 2010).

3.3. Objectives (Why?)

There are many objectives in LA according to the particular point of view of the different stakeholders. Possible objectives of LA include monitoring, analysis, prediction, intervention, tutoring/mentoring, assessment, feedback, adaptation, personalization, recommendation, awareness, and reflection.

3.3.1. Monitoring and Analysis

In monitoring, the objectives are to track student activities and generate reports in order to support decision-making by the teacher or the educational institution. Monitoring is also related to instructional design and refers to the evaluation of the learning process by the teacher with the purpose of continuously improving the learning environment. Examining how students use a learning system and analyzing student accomplishments can help teachers detect patterns and make decision on the future design of the learning activity.

3.3.2. Prediction and Intervention

In prediction, the goal is to develop a model that attempts to predict learner knowledge and future performance, based on his or her current activities and performance. This predictive model can then be used to provide proactive intervention for students who may need additional assistance. The effective analysis and prediction of the learner performance can support the teacher or the educational institution in intervention by suggesting actions that should be taken to help learners improve their performance.

3.3.3. Tutoring and Mentoring

Tutoring is mainly concerned with helping students with their learning (assignments). It is often very domain-specific and limited to the context of a course. A tutor, for example, supports learners in their orientation and introduction into new learning modules as well as instructions of specific subject matter areas within a course. In tutoring processes the control is with the tutor and the focus is rather on the teaching process. In contrast, mentoring goes beyond tutoring and focuses on supporting the learner throughout the whole process – ideally throughout the whole life, but in reality limited to the time that both mentor and learner are part of the same organization. As part of this support, mentors provide guidance in career planning, supervise goal achievement, help preparing new challenges, etc. In mentoring processes the control lies rather with the learners and the focus is on the learning process.

3.3.4. Assessment and Feedback

The objective is to support the (self-)assessment of improved efficiency and effectiveness of the learning process. It is also important to get intelligent feedback to both students and teachers/mentors. Intelligent feedback provides interesting information generated based on data about the user's interests and the learning context.

3.3.5. Adaptation

Adaptation is triggered by the teacher/tutoring system or the educational institution. Here, the goal of LA is to tell learners what to do next by adaptively organizing learning resources and instructional activities according to the needs of the individual learner.

3.3.6. Personalization and Recommendation

In personalization, LA is highly learner-centric, focusing on how to help learners decide on their own learning and continuously shape their PLEs to achieve their learning goals. A PLE-driven approach to learning suggests a shift in emphasis from a knowledge-push to a knowledge-pull learning model (Chatti, 2010). In a learning model based on knowledge-push, the information flow is directed by the institution/teacher. In a learning model driven by knowledge-pull, the learner navigates toward knowledge. One concern with knowledge-pull approaches, though, is information overload. It thus becomes crucial to employ some mechanisms to help learners cope with the information overload problem. This is where recommender systems can play a crucial role to foster self-directed learning. The objective of LA in this case is to help learners decide what to do next by recommending to learners explicit knowledge nodes (i.e. learning resources) and tacit knowledge nodes (i.e. people), based on their preferences and activities of other learners with similar preferences.

3.3.7. Awareness and Reflection

LA can be a valuable tool to support awareness and promote reflection. Students and teachers can benefit from data compared within the same course, across classes, or even across institutions to draw conclusions and (self-)reflect on the effectiveness of their learning or teaching practice. Learning by reflection (or reflective learning) offers the chance of learning by returning to and evaluating past work and personal experiences in order to improve future experiences and promote continuous learning (Boud et al., 1985).

3.4. Methods (How?)

LA applies different methods and techniques to detect interesting patterns hidden in educational data sets. In this section, we describe four techniques that have received particular attention in the LA literature in the last couple of years, namely statistics, information visualization (IV), data mining (DM), and social network analysis (SNA).

3.4.1. Statistics

Most existing learning management systems implement reporting tools that provide basic statistics of the students' interaction with the system. Examples of usage statistics include time online, total number of visits, number of visits per page, distribution of visits over time, frequency of student's postings/replies, and percentage of material read. These reporting tools often generate simple statistical information such as average, mean, and standard deviation.

3.4.2. Information Visualization (IV)

Statistics in form of reports and tables of data are not always easy to interpret to the educational system users. Representing the results obtained with LA methods in a user-friendly visual form might facilitate the interpretation and the analysis of the educational data. Mazza (2009) stresses that thanks to our visual perception ability, a visual representation is often more effective than plain text or data. Different IV techniques (e.g. charts, scatterplot, 3D representations, maps) can be used to represent the information in a clear and understandable format (Romero and Ventura, 2007). The difficult part here is in defining the representation that effectively achieves the analytics objective (Mazza, 2009). Recognizing the power of visual representations, traditional reports based on tables of data are increasingly being replaced by dashboards that graphically show different performance indicators.

3.4.3. Data Mining (DM)

Data mining, also called Knowledge Discovery in Databases (KDD), is defined as "the process of discovering useful patterns or knowledge from data sources, e.g., databases, texts, images, the Web" (Liu, 2006, p. 6). Broadly, data mining methods – which are quite

prominent in the EDM literature – fall into the following general categories: supervised learning (or classification and prediction), unsupervised learning (or clustering), and association rule mining (Han and Kamber, 2006; Liu, 2006).

3.4.4. Social Network Analysis (SNA)

As social networks become important to support networked learning, tools that enable to manage, visualize, and analyze these networks are gaining popularity. By representing a social network visually, interesting connections could be seen and explored in a user-friendly form. To achieve this, social network analysis (SNA) methods have been applied in different LA tasks. SNA is the quantitative study of the relationships between individuals or organizations. In SNA, a social network is modeled by a graph $G = (V, E)$, where V is the set of nodes (also known as vertices) representing actors, and E is a set of edges (also known as arcs, links, or ties), representing a certain type of linkage between actors. By quantifying social structures, we can determine the most important nodes in the network. One of the key characteristics of networks is centrality, which relates to the structural position of a node within a network and details the prominence of a node and the nature of its relation to the rest of the network. Three centrality measures are widely used in SNA: degree, closeness, and betweenness centrality (Wasserman & Faust, 1994).

4. Challenges and Future Research Directions

Significant research has been conducted in LA. Driven by the demands of the new learning environments, there is, however, still a great deal of research that can be done in this area. In light of the LA reference model discussed in the previous section, we discuss various challenges and give insights into potential next steps for LA research.

A particularly rich area for future research is open learning analytics. Open learning analytics has the potential to deal with the challenges in increasingly complex and fast-changing learning environments. Its primary goal is to provide understanding into how learners learn in open and networked learning environments and how educators, institutions, and researchers can best support this process. Open learning analytics refers to an ongoing analytics process that encompasses diversity at all four dimensions of the reference model:

- **What?** It accommodates the considerable variety in learning data and contexts. This includes data coming from traditional education settings (e.g. LMS) and from more open-ended and less formal learning settings (e.g. PLEs, MOOCs).
- **Who?** It serves different stakeholders with very diverse interests and needs.
- **Why?** It meets different objectives according to the particular point of view of the different stakeholders.
- **How?** It leverages a plethora of statistical, visual, and computational tools, methods, and methodologies to manage large datasets and process them into metrics which can be used to understand and optimize learning and the environments in which it occurs.

Open learning analytics is a highly challenging task. It raises a series of challenges and implications for LA practitioners, developers, and researchers. Under the broad umbrella of open learning analytics, we discuss in the next sections various directions that offer new routes for LA research in the future. These research directions and their relation to the four dimensions of the LA reference model are depicted in Figure 2.

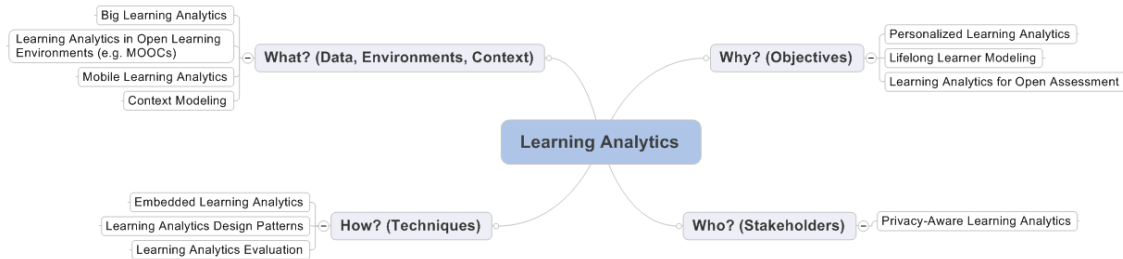


Figure 2. Future Research Directions in Learning Analytics

4.1. Big Learning Analytics

The abundance of educational data, as pointed out in the “what?” dimension of the LA reference model, and the recent attention on the potentiality of efficient infrastructures for capturing and processing large amounts of data, known as big data, have resulted in a growing interest in big learning analytics among LA researchers and practitioners (Dawson et al., 2014). Big learning analytics refers to leveraging big data analytics methods to generate value in TEL environments.

Harnessing big data in the TEL domain has enormous potential. LA stakeholders have access to a massive volume of data from learners’ activities across various learning environments which, through the use of big data analytics methods, can be used to develop a greater understanding of the learning experiences and processes in the new open, networked, and increasingly complex learning environments.

A key challenge in big learning analytics is how to aggregate and integrate raw data from multiple, heterogeneous sources, often available in different formats, to create a useful educational data set that reflects the distributed activities of the learner; thus leading to more precise and solid LA results. Furthermore, handling of big data is a technical challenge because efficient analytics methods and tools have to be implemented to deliver meaningful results without too much delay, so that stakeholders have the opportunity to act on newly gained information in time. Strategies and best practices on harnessing big data in TEL have to be found and shared by the LA research community.

4.2. Learning Analytics in Open Learning Environments (e.g. MOOCs)

As stressed in the “what?” dimension of the LA reference model, learning is increasingly taking place in open and networked learning environments. In the past few years, Massive Open Online Courses (MOOCs) have gained popularity as a new form of open learning. MOOCs can be roughly classified in two groups. On the one hand there are xMOOCs.

Although they gained a lot of attention they can be seen as a replication of traditional learning management systems (LMS) at a larger scale. Still they are closed, centralized, structured, and teacher-centered courses that emphasize video lectures and assignments. In xMOOCs all services available are predetermined and offered within the platform itself. On the other hand there is the contrasting idea of cMOOCs combining MOOCs with the concept of personal learning environment (PLE). In contrast to xMOOCs, cMOOCs are open-ended, distributed, networked, and learner-directed learning environments where the learning services are not predetermined, and most activities take place outside the platform (Yousef et al., 2014a).

MOOC environments provide an exciting opportunity for the LA community. They capture and store large data sets from learners' activities that can provide insight into the learning processes. MOOCs house a wide range of participants with diverse interests and needs. Understanding how learners are engaging with the MOOC and analyzing their activities in the course is a first step towards more effective MOOCs. Applying LA in MOOCs can thus provide course providers and platform developers with insights for designing effective learning environments that best meet the needs of MOOC participants.

In addition, due to the open nature of MOOCs, learners are often overwhelmed with the flow of information in MOOCs. Moreover, in MOOCs it is difficult to provide personal feedback to a massive number of learners. LA can provide great support to learners in their MOOC experience. LA that focuses on the perspectives of learners can help to form the basis for increased learner awareness and effective personalized learning.

Recognizing these opportunities, several MOOC studies recommended to apply LA tools to monitor the learning process, identify difficulties, discover learning patterns, provide feedback, and support learners in reflecting on their own learning experience (Fournier et al., 2011; Yousef et al., 2014b). Despite the wide agreement that LA can provide value to different MOOC stakeholders, the application of LA techniques on MOOC datasets is rather limited until now. Most of the LA implementations in MOOCs so far meet the needs of the course providers and have primarily focused on addressing low completion rates, investigating learning patterns, and supporting intervention (Clow, 2013b; Mirriahi & Dawson, 2013). A focus on the perspectives of learners in MOOCs has the potential to extend LA beyond completion rates and interventions to support personalization, feedback, assessment, recommendation, awareness, and self-reflection.

As compared to xMOOCs, cMOOCs offer particular challenges for LA practice. In fact, in a cMOOC, participation is emergent, fragmented, diffuse, and diverse (Clow, 2013b; McAulay et al., 2010) and most of the learning activities happen outside the platform across different learning environments. Operating in cMOOCs require a shift towards LA on more challenging data sets across a variety of different sites with different standards, owners, and levels of access (Ferguson, 2012; Fournier et al., 2011).

4.3. Mobile Learning Analytics

Another important research direction related to the "what?" dimension of the LA reference model is mobile learning analytics. The widespread use of mobile technologies has led to an increasing interest in mobile learning (Thüs et al., 2012). Mobile technologies open an unprecedented opportunity to conduct research on mobile learning analytics. Mobile

learning analytics refers to LA mediated through mobile and ubiquitous technologies. It includes harnessing the data of mobile learners and presenting the analytics results on mobile devices. Mobile learning analytics bears a number of challenges. It has to handle the specificity of the mobile devices with respect to their constraints. A mobile device has only a specific amount of battery power which discharges very quickly when network communication is extensively used. The limited bandwidth is also one gap which has to be considered seriously, as huge amounts of data required for a LA task cannot be transferred so easily. Data has to be transferred in real time as the usage intervals of the mobile learning platform are frequent but very short, thus the user cannot be kept waiting for a long period of time. Furthermore, a mobile device is not as powerful as a personal computer. Therefore, time-consuming computing operations have to be done on external servers or by other technologies which provide enough computation power to solve hard computational tasks. The mobile device should only process the presentation of the analytics results. The possibility to display graphical elements (e.g. dashboards) on the screen depends on its resolution and size. Furthermore, mobile learning is normally done on-the-fly, this means it is very fast and mostly during spare time. Thus, the provided analytics results have to be personalized to the needs of the user. It is also important to leverage the context attribute in mobile learning in order to give learners the support they need when, how, and where they need it.

4.4. Context Modeling

In addition to data and environments, the “what?” dimension of the LA reference model refers to the context in which learning occurs. In fact, learning is happening in a world without boundaries. Learners are collaborating more than ever beyond classroom boundaries which become more and more irrelevant. Thus, for data gathering, the context in which events take place plays a pivotal role.

Context is crucial in mobile learning environments. In fact, a major benefit of mobile devices is that they enable learning across contexts. Harnessing context in a mobile learning experience has a wide range of benefits including personalization, adaptation, intelligent feedback, and recommendation. Adding context information to a set of mobile applications may lead to presenting more personalized data to the learner and to more motivating mobile learning experiences.

Context is also a key factor in workplace learning, self-directed learning, and lifelong learning to achieve effective learning activities. Nowadays, knowledge workers and self-directed lifelong learners perform in highly complex knowledge environments. They have to deal with a wide range of activities they have to manage every day. And, they have to combine learning activities and their private and professional daily life. The challenge here is the linking of formal and informal learning activities and the bridging of learning activities across different contexts.

To date, there is little support for learners that typically try to learn in different contexts and must align their learning activities to their daily routines. LA tools that are smoothly integrated into learners' daily life can offer solutions to address this challenge. There is a need to focus on the context attribute in LA by developing methods for effective context capturing and modeling. A context model should reflect a complete picture of the learner's context information.

4.5. Privacy-Aware Learning Analytics

As stated in the “who?” dimension of the LA reference model, the application of LA can be oriented toward different stakeholders. In the internet era, keeping personal information private is a big challenge. This challenge transfers to LA as well. There are different approaches and research is being done in order to understand and tackle the ethical and privacy issues that arise with practical use of LA in the student context. Pardo and Siemens (2014), for instance, provide four practical principles that researchers should consider about privacy when working on an LA tool. These practical principles are transparency, student control over the data, security, and accountability and assessment. Slade and Prinsloo (2013) propose an ethical framework with six guiding principles for privacy-aware LA implementations. These include (1) learning analytics as moral practice, (2) students as agents, (3) student identity and performance are temporal dynamic constructs, (4) student success is a complex and multidimensional phenomenon, (5) transparency, and (6) higher education cannot afford to not use data. Some of the guiding principles are overlapping, or the same as the principles suggested by Pardo and Siemens (2014). It is crucial to embrace these principles while creating new opportunities for analytics-supported learning.

In general, we have a vast range of possibilities for LA research, but we should always keep in hindsight the ethical and privacy challenges of the stakeholders while trying to institutionalize LA and include it into the day-to-day learning activities. Further research is needed to investigate approaches that can help design LA solutions where ethical and privacy issues are considered.

4.6. Personalized Learning Analytics (Goal-oriented Learning Analytics)

In relation to the “why?” dimension in the LA reference model, the various objectives in LA (e.g. monitoring, analysis, prediction, intervention, tutoring, mentoring, assessment, feedback, adaptation, personalization, recommendation, awareness, reflection) need a tailored set of indicators and metrics to serve different stakeholders with very diverse needs. Current implementations of LA rely on a predefined set of indicators and metrics. This raises questions about what sort of indicators and metrics are helpful in each learning situation. Ideally, LA tools should support an exploratory, real-time user experience that enables data exploration and visualization manipulation based on individual interests of users. The challenge is thus to define the right Goal / Indicator / Metric (GIM) triple before starting the LA exercise. Giving users the opportunity to interact with the LA tool by defining their goal and specifying the indicator/metric to be applied is a crucial step for more effective analytics results. This would also make the analytics process more transparent, enabling users to see what kind of data is being used and for which purpose. From a technical perspective, templates and rule engines can be used to achieve goal-oriented LA for improved personalized learning.

4.7. Lifelong Learner Modeling

The “why?” dimension of the LA reference model in particular encompasses learner-centered LA objectives, such as intelligent feedback, adaptation, personalization, and recommendation. In order to achieve these objectives, learner modeling is a crucial task. A learner (user) model is a representation of information about an individual learner that is essential for adaptation and personalization tasks. The six most popular and useful features in learner modeling include the learner’s knowledge, interests, goals, background, individual traits, and context (Brusilovsky and Millan, 2007). The challenge is to create a thorough “polycontextual” learner model that can be used to trigger effective intervention, adaptation, personalization, or recommendation actions. This is a highly challenging task since learner activities are often distributed over open and increasingly complex learning environments. A big challenge to tackle here is lifelong learner modeling. Kay and Kummerfeld (2011) define a lifelong learner model as a store for the collection of learning data about an individual learner. The authors note that to be useful, a lifelong learner model should be able to hold many forms of learning data from diverse sources and to make that information available in a suitable form to support learning. The aim is that data gathered from different learning environments would be fed into a personal lifelong learner model, which would be the location where the learner can archive all learning activities throughout her life. This model must store very heterogeneous data. It would also contain both long-term and short-term goals of the learner.

The capacity to build a detailed picture of the learner across a broader learning context beyond the classroom would provide more accurate analytics results. Future research in this area should focus on effective and efficient LA methods to mine, aggregate, manage, share, reuse, visualize, interpret, and leverage learner models in order to better induce and support personalized and lifelong learning. Thereby, several issues have to be taken into account, including questions about privacy, scalability, integration, and interoperability.

4.8. Learning Analytics for Open Assessment

In relation to the “why?” dimension in the LA reference model, assessment constitutes a core component of the learning process and one important objective for LA. Recent years have seen a growing interest in learner-centered, open, and networked learning models (e.g. PLEs, OER, MOOCs). Unlike assessment in classroom settings, the methodology to assess learning in open environments has not yet been well investigated. The move to open learning models raises challenges with regard to the assessment of learning. Learner-centered and networked learning requires new assessment models that address how to recognize and evaluate self-directed, network, and lifelong learning achievements. We need thus to think about an innovative open assessment framework for accrediting and recognizing open learning in an efficient, valid, and reliable way.

Open assessment is an all-encompassing term combining different assessment methods for recognizing learning in open and networked learning environments. In open assessment, summative, formative, formal, informal, peer, network, self-, and e-assessment converge to allow lifelong learners to gain recognition of their learning. It is an

agile way of assessment where anyone, anytime, anywhere, can participate towards the assessment goal. Open assessment is an ongoing process across time, locations, and devices where everyone can be assessor and assessee.

Open assessment is a major goal for LA. However, harnessing LA for assessment purposes has not yet been well investigated. It is important to implement LA-based open assessment methods throughout the entire learning process that enable learners to continuously receive feedback, validation, and recognition of their learning, thus motivating and engaging them in their learning journey. LA can support open assessment in various ways. These include:

- Improving e-assessment (e.g. automatic or semi-automatic scoring, automatic evaluation of free text answers, automatic issuing of badges).
- Integrating summative assessment (e.g. score and rank) to support prediction and intervention.
- Generating real-time predictions of learner performance.
- Helping teachers/mentors give accurate and personalized feedback.
- Visualizing the learning progress and achievement to support awareness, trigger self-reflection, and promote self-assessment.
- Generating quality feedback to learners to help them enhance their learning practice. Continuous feedback is important for effective learning. It facilitates awareness, reflection, and self-assessment during the learning process and reinforces learner's self-esteem and motivation.
- Providing an explanation on how and why a feedback/assessment has been given.
- Gauging strengths and weaknesses in learner participation in learning activities in collaborative learning environments.
- Validating peer assessment, which can be prone to biases.

4.9. Embedded Learning Analytics

The “how?” dimension of the LA reference model suggests different techniques and tools that can be applied in LA. Integration is a challenge that has to be taken into account in the development of LA tools. In fact, LA is most effective when it is an integrated part of the learning environment. Hence, integration of LA into the learning practice of the different stakeholders is important and should be accompanied by appropriate didactical and organizational frameworks. LA developers will have to provide tools, which should optimally be embedded into the standard TEL toolsets of the learners and teachers.

Moreover, effective LA tools are those, which minimize the time frame between analysis and action. Thus, it is crucial to follow an embedded LA approach by implementing contextualized LA applications that give useful information at the right place and time; thus getting away from traditional dashboard-based LA.

4.10. Learning Analytics Design Patterns

Learning analytics design patterns constitute a further research direction related the “how?” dimension of the LA reference model. For the development of usable and useful LA tools, guidelines and design patterns should be taken into account. Appropriate visualizations could make a significant contribution to understanding the large amounts of educational data. Statistical, filtering, and mining tools should be designed in a way that can help learners, teachers, and institutions to achieve their analytics objectives without the need for having an extensive knowledge of the techniques underlying these tools. In particular, educational data mining tools should be designed for non-expert users in data mining. Moreover, performance, scalability, and extensibility should be taken into consideration in order to allow for incremental extension of data volume and analytics functionality after a system has been deployed. Also, student and teacher involvement is the key to a wider user acceptance, which is required if LA tools are to serve the intended objective of improving learning and teaching. The number and hierarchy of stakeholders bear conflicts. Involving all stakeholders and supporting all their interests is a complex task that has to be solved, since some of these interests might be contradictory. For instance, the usage of LA by administration staff for finding best practice examples of technology-enhanced teaching could offend teachers because they might feel assessed and controlled. The same could be true for students who might fear that personal data will not be used for formative evaluation but for assessment and grading. This could lead to the unintended effect that teachers and/or students are not motivated to use new technologies and participate in analytics-based TEL scenarios.

4.11. Learning Analytics Evaluation

Another research opportunity related the “how?” dimension of the LA reference model is the effective evaluation of the developed LA tools. LA tools (as most interfaces) are generative artifacts: They do not have value in themselves, but they produce results in a particular context. This means that an LA tool is used by a particular user (e.g. student, teacher, mentor, researcher), on a particular data set, and for a particular reason. Hence, the evaluation of such tool is very complicated and diverse. Ellis and Dix (2006) argue that for search of validation of generative artefacts, empirical evaluation is methodologically unsound. Any empirical evaluation, cannot tell you, in itself, that the LA tool works, or does not work. Dyckhoff et al. (2013) summarize the goals of LA evaluation in three major goal groups: (a) explicitly inform the design of LA tools, (b) involve a behavioral reaction of the teacher, and (c) involve a behavioral reaction of the student. The authors inquire how does an LA tool, which has been carefully designed according to the mentioned goals and requirements, influence the behavior of its users, and therefore how does this tool influence practical learning situations? In their research they found only few publications that reported findings related to the behavioral reaction of teachers and students while using LA tools. They strongly call for impact analysis and usefulness evaluation of the LA tools.

Effective evaluation is difficult and problem-prone, but it is essential to support the LA task. There are several aspects that one needs to take into account in order to do an effective evaluation. The first thing one has to consider is the purpose, and the gains from the evaluation. The evaluators need to carefully design the goals and try to meet them with the

tool's evaluation. Once the evaluators have set their goals for the evaluation, they need to think about the measures and tasks that are going to be encompassed in the evaluation. It is important to define the appropriate indicators and metrics. The evaluators also need to consider collecting qualitative data and use qualitative methods in pair with the quantitative evaluation (wherever possible). A mixed-method evaluation approach that combines both quantitative and qualitative evaluation methods can be very powerful in order to capture when, how often, and why a peculiar behavior happens in the learning environment.

How can we evaluate the usability, usefulness, and learning impact of LA? While usability is relatively easy to evaluate, the challenge is to investigate how LA could impact learning and how this could be evaluated. Measuring the impact of LA tools is a challenging task, as the process needs long periods of time as well as a lot of effort and active participation from researchers and participants. Moreover, the analysis of the qualitative data from the evaluation is always prone to personal interpretations and biased while making conclusions (Dyckhoff et al., 2013). Thus, further research is required to investigate effective mixed-method evaluation approaches that do not only focus on the usability of the LA tool, but also aim at measuring its impact on learning.

5. Conclusions

In the last few years, there has been a growing interest in learning analytics (LA) in technology-enhanced learning (TEL). LA approaches share a movement from data to analysis to action to learning. The TEL landscape is changing, and this should be reflected in future LA approaches for more effective learning experiences. In this paper, we gave an overview of the range of possibilities opened up by LA. We discussed a reference model that provides a view of LA and its related concepts based on four dimensions: data, environments, context (what?), stakeholders (who?), objectives (why?), and methods (how?). In light of the LA reference model, we suggested several promising directions for future LA research. These include big LA, LA in open learning environments, mobile LA, and context modeling (related to the “what?” dimension), privacy-aware LA (related to the “who?” dimension), personalized LA, lifelong learner modeling, and LA for open assessment (related to the “why?” dimension), as well as embedded LA, LA design patterns, and LA evaluation (related to the “how?” dimension). These research directions require a number of challenges to be addressed in order to capture the full potential of LA. These challenges include handling increasing data volume, heterogeneity, fragmentation, interoperability, integration, performance, scalability, extensibility, real-time operation, reliability, usability, finding meaningful indicators/metrics, using mixed-method evaluation approaches, and most important dealing with data privacy issues. This paper makes a significant contribution to LA research because it provides a more grounded view of LA and its related concepts and identifies a series of challenges and promising research directions in this emerging field, which have been lacking until now.

References

Boud, D.; Keogh, R.; Walker, D.: Reflection: Turning Experience into Learning. In: *Promoting Reflection in Learning: a Model*. Routledge Falmer, New York, 1985, pp. 18-40.

Brusilovsky, P.; Millan, E.: User Models for Adaptive Hypermedia and Adaptive Educational Systems. In: Brusilovsky, P.; Kobsa, A.; Nejdil, W. (Eds.): *The Adaptive Web*, LNCS 4321. Springer, Berlin Heidelberg, 2007, pp. 3–53.

Campbell, J.P.; DeBlois, P.B.; Oblinger, D.G.: Academic Analytics: A New Tool for A New Area. In: *EDUCAUSE Review*, July/August 2007, pp. 41-57.

Chatti, M.A.; Dyckhoff, A.L.; Schroeder, U.; Thüs, H. (2012a): A reference model for learning analytics. In: *International Journal of Technology Enhanced Learning (IJTEL)*, 4(5/6), 2012, pp. 318-331.

Chatti, M. A.; Schroeder, U.; Jarke, M. (2012b): LaaN: Convergence of Knowledge Management and Technology-Enhanced Learning. In: *IEEE Transactions on Learning Technologies*, 5(2), 2012, pp. 177–189.

Chatti, M. A.: *The LaaN Theory*. In: *Personalization in Technology Enhanced Learning: A Social Software Perspective*. Aachen, Germany: Shaker Verlag, 2010, pp. 19-42. <http://mohamedaminechatti.blogspot.de/2013/01/the-laan-theory.html> (last check 2014-11-03)

Clow, D. (2013a): An overview of learning analytics. In: *Teaching in Higher Education*, 18(6), 2013, pp. 683–695.

Clow, D. (2013b): MOOCs and the Funnel of Participation. In: *Proceedings of the Third International Conference on Learning Analytics & Knowledge*. ACM New York, NY, USA, 2013, pp. 185-189.

Dawson, S.; Gašević, D.; Siemens, G.; Joksimovic, S.: Current state and future trends: a citation network analysis of the learning analytics field. In: *Proceedings of the Fourth International Conference on Learning Analytics & Knowledge*. ACM New York, NY, USA, 2014, pp. 231-240.

Dyckhoff, A.; Lukarov, V.; Muslim, A.; Chatti, M. A.; Schroeder, U.: Supporting action research with learning analytics. In: *Proceedings of the Third International Conference on Learning Analytics & Knowledge*. ACM New York, NY, USA, 2013, pp. 220-229

EDUCAUSE (2010): 7 Things you should know about analytics. 2010. <http://www.educause.edu/ir/library/pdf/ELI7059.pdf> (last check 2014-11-03)

Ellis, G.; Dix, A.: An explorative analysis of user evaluation studies in information visualisation. In: *Proceedings of the 2006 Conference on Beyond Time and Errors: Novel Evaluation Methods For information Visualization*. ACM Press, New York, NY, 2006, pp. 1-7.

Ferguson, R.: Learning analytics: drivers, developments and challenges. In: *International Journal of Technology Enhanced Learning*, 4(5/6), 2012, pp. 304-317.

Fournier, H.; Kop, R.; Sitlia, H.: The value of learning analytics to networked learning on a personal learning environment. *Proceedings of the LAK '11 Conference on Learning Analytics and Knowledge*, 2011, pp. 104-109.

Goldstein, P.J.; Katz, R.N.: *Academic Analytics: The Uses of Management Information and Technology in Higher Education*. ECAR Research Study, Vol. 8. , 2005. <http://www.educause.edu/ECAR/AcademicAnalyticsTheUsesofMana/158588> (last check 2014-11-03)

Han, J.; Kamber, M.: *Data Mining: Concepts and Techniques*. Elsevier, San Francisco, CA, 2006.

Kay, J.; Kummerfeld, B.: *Lifelong learner modeling*. In: Durlach, P.J.; Lesgold, A.M. (Eds.): *Adaptive Technologies for Training and Education*. Cambridge University Press, Cambridge, 2011, pp. 140-164.

Norris, D.; Baer, L.; Leonard, J.; Pugliese, L.; Lefrere, P.: *Action Analytics: Measuring and Improving Performance That Matters in Higher Education*. In: *EDUCAUSE Review*, 43(1), 2008, pp. 42-67.

Liu, B.: *Web Data Mining*. Springer, Berlin, Heidelberg, 2006.

Mazza, R.: *Introduction to Information Visualization*. Springer-Verlag, London, 2009.

McAulay, A.; Stewart, B.; Siemens, G.; Cormier, D.: *The MOOC model for digital practice*. 2010. http://www.elearnspace.org/Articles/MOOC_Final.pdf (last check 2014-11-03)

Mirriahi, N.; Dawson, S.: *The Pairing of Lecture Recording Data with Assessment Scores: A Method of Discovering Pedagogical Impact*. In: *Proceedings of the Third International Conference on Learning Analytics & Knowledge*. ACM New York, NY, USA, 2013, pp. 180-184.

Pardo, A.; Siemens, G.: *Ethical and Privacy Principles for Learning Analytics*. In: *British Journal of Educational Technology*, 45(3), 2014, pp. 438-450.

Romero, C.; Ventura, S.: *Educational Data Mining: a Survey from 1995 to 2005*. In: *Expert Systems with Applications*, 33(1), 2007, pp. 135-146.

Siemens, G.; Baker, S.J.: *Learning analytics and educational data mining: Towards communication and collaboration*. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK'12)*, Vancouver, Canada, 29th April–2nd May 2012. ACM, New York, NY, 2012, pp. 252–254.

Siemens, G.; Long, P.: *Penetrating the Fog: Analytics in Learning and Education*. In: *EDUCAUSE Review*, 46(5), September/October 2011.

Slade, S.; Prinsloo, P.: *Learning analytics: ethical issues and dilemmas*. In: *American Behavioral Scientist*, 57(10), 2013, pp. 1509–1528.

Thüs, H.; Chatti, M. A.; Yalcin, E.; Pallasch, C.; Kyrlyiuk, B.; Mageramov, T.; Schroeder, U.: *Mobile Learning in Context*. In: *International Journal of Technology Enhanced Learning*, 4(5/6), 2012, pp. 332–344.

Wasserman, S.; Faust, K.: *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, 1994.

Yousef, A. M. F.; Chatti, M. A.; Schroeder, U.; Wosnitza, M.; Jakobs, H. (2014a): *MOOCs - A Review of the State-of-the-Art*. In: *Proceedings of the CSEDU 2014 conference*, Vol. 3, 2014, pp. 9-20.

Yousef, A. M. F.; Chatti, M. A.; Schroeder, U.; Wosnitza, M. (2014b): *What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs*. In: *Proceedings of the 14th IEEE International Conference on Advanced Learning Technologies*, 2014, pp. 44-48.