A Reference Model for Learning Analytics

M.A. Chatti, A.L. Dyckhoff, U. Schroeder, and H. Thüs
Informatik 9 (Learning Technologies), RWTH Aachen University
{chatti,dyckhoff,Schroeder,thues}@informatik.rwth-aachen.de

ABSTRACT

Recently, there is an increasing interest in learning analytics in Technology Enhanced Learning (TEL). Generally, learning analytics deals with the development of methods that harness educational data sets to support the learning process. Learning analytics (LA) is a multi-disciplinary field involving machine learning, artificial intelligence, information retrieval, statistics, and visualization. LA is also a field in which several related areas of research in TEL converge. These include academic analytics, action research, educational data mining, recommender systems, and personalized adaptive learning. In this paper, we investigate the connections between LA and these related fields. We describe a reference model for LA based on four dimensions, namely data and environments (what?), stakeholders (who?), objectives (why?), and methods (how?). We then review recent publications on LA and its related fields and map them to the four dimensions of the reference model. Furthermore, we identify various challenges and research opportunities in the area of LA in relation to each dimension.

Keywords

Learning analytics, Academic analytics, Educational data mining, Action research, Literature review, Reference model

1. **Introduction**

In the last few years, there has been a growing interest in the automatic analysis of educational data to enhance the learning experience, a research area referred to recently as learning analytics. The 2011 Horizon Report identified learning analytics as a possible key future trend in learning and teaching (Johnson et al., 2011). Learning analytics is not a genuine new research area. It actually borrows from different related fields and synthesizes several existing techniques. The connections between learning analytics and its related areas of research (i.e. academic analytics, action research, educational data mining, recommender systems, and personalized adaptive learning) are not well addressed in the emerging learning analytics literature. In this paper, we investigate the evolution of learning analytics in recent years and discuss a reference model that enables to classify the literature in learning analytics. This model can also foster a common understanding of key concepts in this emerging field.

The remainder of this paper is structured as follows. In section 2, we investigate the connections between learning analytics and its related fields. Section 3 highlights the main steps of the learning analytics process. In section 4, we discuss a four-dimensional reference model for learning analytics and identify future challenges in relation to each dimension. We then review recent relevant publications around learning analytics in section 5 and categorize them according the four dimensions of the reference model. Finally, section 6 gives a summary of the main results of this paper and highlights directions for future work.

2. Learning Analytics Definitions and Related Concepts

Different definitions have been provided for the term ‘learning analytics’. Learning analytics (LA) is defined on the LAK11 website (https://tekri.athabascau.ca/analytics/) as "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". Elias (2011) describes LA as “an emerging field in which sophisticated analytic tools are used to improve learning and education”. Siemens (2010) views LA as "the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning". LA is defined in EDUCAUSE’s Next Generation learning initiative as “the use of data and models to predict student progress and performance, and the ability to act on that information” (as cited in Siemens, 2010). According to Johnson et al. (2011), LA "refers to the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues". Although different in some details, these 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.

LA concepts and methods are drawn from a variety of related research fields including academic analytics, action research, educational data mining, recommender systems, and personalized adaptive learning. In the next sections, we investigate the connections between LA and these related fields.

2.1. Academic Analytics

The term academic analytics (AA), first used by WebCT, was introduced by Goldstein and Katz (2005) to describe the application of business intelligence tools and practices in higher education (Elias, 2011; Norris et al., 2008a). Business intelligence (also called analytics) is "a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions" (Goldstein and Katz, 2005). In their study, Goldstein and Katz use the term to describe the scope of their study on how academic institutions gather, analyze, and use data to support decision-making. Campbell et al. (2007) as well as Campbell and Oblinger (2007) provide a narrower definition of AA that focuses on student success, defined in student retention and graduation rates. AA "has the potential to create actionable intelligence to improve teaching, learning, and student success" (Campbell et al., 2007). It "marries data with statistical techniques and predictive modeling to help faculty and advisors determine which students may face academic difficulty, allowing interventions to help them succeed (Campbell and Oblinger, 2007). The examples in the AA literature refer mostly to the problem of detecting “at risk students”, that is, those students that might drop out of a course or abandon their studies. An example of the use of AA is the Signals project in use at Purdue University (Arnold, 2010; Tanes et al., 2011).

In general, early AA initiatives since the late 1990s applied analytics methods to meet the needs of educational institutions. They only focused on enrollment management and the prediction of student academic success and were restricted to statistical software (Campbell et al., 2007; Campbell and Oblinger, 2007; Goldstein and Katz, 2005; Norris et al., 2008a). Besides just serving the goals of educational institutions, applications of LA can be oriented toward different stakeholders including learners and teachers. Beyond enrollment 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.

Norris et al. (2008b) use the term action analytics to describe "the emergence of a new generation of tools, solutions, and behaviors that are giving rise to more powerful and effective utilities through which colleges and universities can measure performance and provoke pervasive actions to improve it". According to the authors, action analytics takes AA on the next level. Crucial for the authors is "to move from data to reporting to analysis to action" (Norris et al., 2008a). This view of analytics is, however, still defined at the institutional level and does not target other key stakeholders including learners and teachers.

2.2. Action Research

Action research is a methodology that has become increasingly popular and has been well developed in education, specifically in teaching at universities and schools in many countries around the world (McNiff and Whitehead, 2006). It enables teachers themselves to investigate and evaluate their work (Altrichter et al., 1996). The purpose of educational action research is to improve teaching practice and assure quality. Based on a concrete teaching situation, teachers and students systematically investigate arising research questions, whereas action and reflection correlate. Through iterative cycles of action, perception and evaluation teaching can regularly be matched and adjusted to all learners' needs. Thereby teachers learn more about their teaching and are enabled to improve their personal teaching skills (Hinchey, 2008).

Although the goals behind action research and LA are very similar, a difference can be seen in the initial trigger of related study projects. While action research projects usually start with a research question that arises from teaching practice, LA projects often evolve based on observations made with regard to already collected data. Action research projects, therefore, also often use qualitative methods to generate a holistic picture of the learning situation, while LA are mostly based on quantitative methods. Furthermore, stakeholders in action research projects are mainly teachers and students, whereas LA might address other stakeholders, like system designers or institutional staff, as well.

2.3. Educational Data Mining

Educational data mining (EDM) has emerged as an independent research area in recent years, starting with several EDM workshops which have led to the establishment of an annual International Conference on Educational Data Mining in 2008 and the appearance of the Journal of Educational Data Mining (JEDM) and specific books on EDM, such as 'The Handbook of Educational Data Mining' (Romero et al., 2010a). EDM is concerned with developing methods to explore the unique types of data that come from an educational context and, using these methods, to better understand students and the settings in which they learn (Romero et al., 2010a). From a technical perspective, EDM is the application of data mining techniques to educational data, and so, its objective is to analyze this type of data in order to resolve educational research issues and understand the setting in which students learn (Baker and Yacef, 2009; Barnes et al., 2009). Romero and Ventura (2007), Baker and Yacef (2009), and Romero and Ventura (2010) provide an excellent review of how EDM has developed in recent years as well as the major trends in EDM research up to 2009.

The analysis domain, data, process, and objectives in LA and EDM are quite similar. 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 techniques used for LA can be quite different from those used in EDM. EDM basically focus on the application of typical data mining techniques (i.e. clustering, classification, and association rule mining) to support teachers and students in analyzing the learning process. In addition to data mining techniques, LA further includes other methods, such as statistical and visualization tools or social network analysis (SNA) techniques, and puts them into practice for studying their actual effectiveness on the improvement of teaching and learning.

2.4. Recommender Systems

Generally, recommender systems aggregate data about user's behavior or preferences in order to draw conclusions for recommendation of items she most likely might be interested in. Technically, recommender systems are classified into the following classes, based on how recommendations are made (Adomavicius & Tuzhilin, 2005):

- Content-based recommendations: The user will be recommended items similar to the ones the user preferred in the past;
- Collaborative Filtering (CF): The user will be recommended items that people with similar tastes and preferences liked in the past. In CF, an item is considered as a black box, and user interactions with the item are used to recommend an item of interest to the user. To recommend items to user, we either find similar users and recommend items liked by these users (user-based analysis), or we find items similar to an item of interest (item-based analysis).
- Hybrid approaches: These methods combine collaborative and content-based methods.

Techniques for content-based and CF recommendations include information retrieval methods (e.g. TF-IDF) and various machine learning techniques, such as Bayesian classifiers, decision trees, artificial neural networks, and clustering.

Recommendation is increasingly used in TEL research in the last few years as a core objective of the LA task. And, the techniques used for recommendation (i.e. information retrieval and machine learning algorithms) are also common in LA applications, but there are open research questions of how algorithms and methods have to be adapted and optimized to be transferred from the domain of commercial recommendations.

2.5. Personalized Adaptive Learning

There are many definitions of adaptation in educational systems. The two main terms usually involved are adaptivity and adaptability. Adaptivity is the ability to modify course materials using different parameters and a set of pre-defined rules. Adaptability is the possibility for learners to personalize the course materials by themselves (Burgos, 2007). Early literature on personalized adaptive learning has focused on adaptivity. Most of this literature has focused on adaptive intelligent educational systems, such as Intelligent Tutoring Systems (ITS) and Adaptive Hypermedia Systems (AHS). A common idea behind those adaptive educational systems is that, based on the information about the learner and the current context, an appropriate adaptation method should be chosen to adapt the presentation of the course material to the individual learner. More recent literature on personalized adaptive learning criticized that traditional approaches are very much top-down and ignore the crucial role of the learners in the learning process. A new vision of personalized learning has been introduced in the last few years based on the concept of Personal Learning Environments (PLE). In PLEs, personalization is triggered by the learner, rather than processed by an intelligent system. The PLE-driven approach to learning does not focus on adaptivity, i.e. the top-down configuration of a set of learning elements following some pre-defined rules. It rather focuses on adaptability and emphasizes learner's self-direction, self-organization, and self-control. Consequently, learners are not responsible for adapting themselves to the requirements of the institution or the instructor. They are rather responsible for creating and maintaining their very own learning environments, self-adapted to their individual needs. From this perspective, personalization can be defined as the ability on the part of the learner to learn the way she deems fit (Chatti, 2010).

Most personalized adaptive learning solutions applied potential LA tools either to tell learners what to do next by automatically matching teaching material to the individual needs of the learner (i.e. adaptivity) or to help learners decide what to do next by recommending them different learning entities, based on their preferences (i.e. adaptability).

In personalized adaptive learning as well as in LA, learner profiling is a crucial task. Both fields focus on learner modeling as the core for achieving adaptive and personalized learning environments, which will be able to take into account the heterogeneous needs of learners and provide them with tailored learning experience suited for their unique needs.

2.6. Learning Analytics

LA builds upon the above-mentioned research areas. In this paper, we view LA as a generic all-encompassing term to describe a 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.

3. Learning Analytics Process

As illustrated in Figure 1, the overall LA process is often an iterative cycle and is generally carried out in three major steps: (1) data collection and pre-processing, (2) analytics and action, and (3) post-processing.

Data collection and pre-processing: Educational data is the foundation of the LA process. The first step in any LA effort is to collect data from various educational environments and systems. This step is critical to the successful discovery of useful patterns from the data. The collected data may be too large and/or involve many irrelevant attributes, which call for data pre-processing (also referred to as data preparation) (Liu, 2006). Data pre-processing also allows transforming the data into a suitable format that can be used as input for a particular LA method. Several data pre-processing tasks, borrowed from the data mining field, can be used in this step. These include data cleaning, data integration, data transformation, data reduction, data modeling, user and session identification, and path completion (Han and Kamber, 2006, Liu, 2006; Romero & Ventura, 2007)

Analytics and action: Based on the pre-processed data and the objective of the analytics exercise, different LA techniques can be applied to explore the data in order to discover hidden patterns that can help to provide a more effective learning experience. The analytics step does not only include the analysis and visualization of information, but also actions on that information. Taking actions is the primary aim of the whole analytics process. These actions include monitoring, analysis, prediction, intervention, assessment, adaptation, personalization, recommendation, and reflection. We discuss these actions in details in section 4.3.

Post-processing: For the continuous improvement of the analytics exercise post-processing is fundamental. It can involve compiling new data from additional data sources, refining the data set, determining new attributes required for the new iteration, identifying new indicators/metrics, modifying the variables of analysis, or choosing a new analytics method.

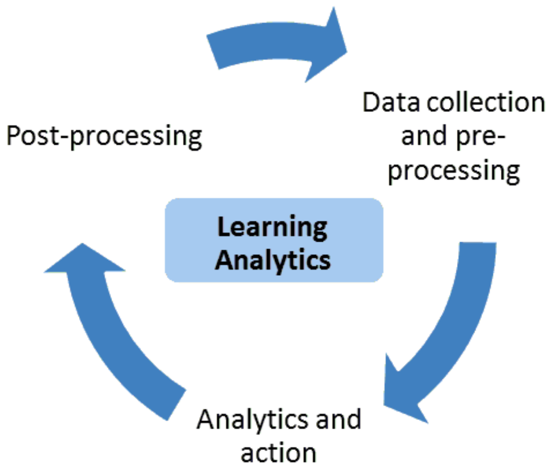

Figure 1. Learning Analytics Process

4. A Reference Model for Learning Analytics

In the following sections, we describe a reference model for LA based on four dimensions and identify various challenges and research opportunities in the area of LA in relation to each dimension.

As depicted in Figure 2, 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?

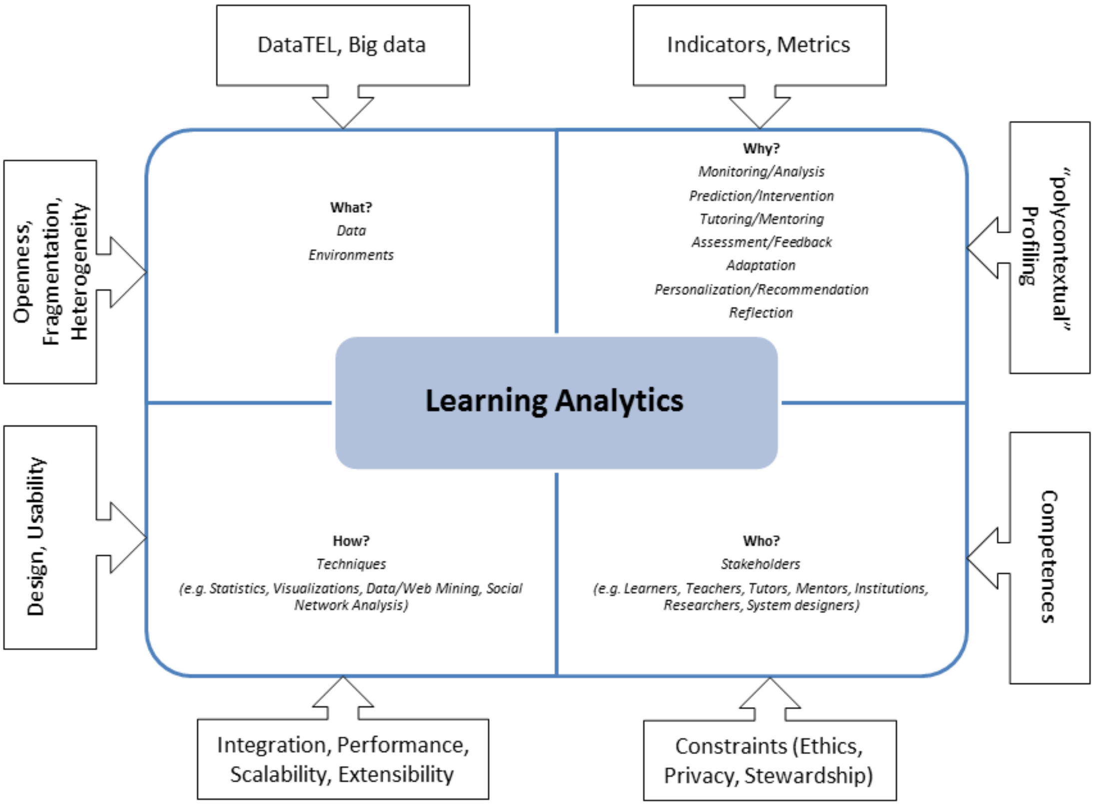

Figure 2. Learning Analytics Reference Model

4.1. Data and Environments (What?)

LA is a data-driven approach. 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). These include commercial LMS, e.g., Blackboard, own developments, e.g., L²P, or open source LMS, e.g., Moodle. 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 distant 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. 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. "Learning and knowledge creation is often distributed across multiple media and sites in networked environments", so that "traces of such activity may be fragmented across multiple logs and may not match analytic needs" (Suthers and Rosen, 2011).

As learning tools and resources are increasingly moving into the cloud, the challenge 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 how to deal with the data volume have to be found and shared by the LA research community.

4.2. 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; Campbell and Oblinger, 2007), develop student recruitment policies, adjust course planning, determine hiring needs, or make financial decisions (Educause, 2010).

LA researchers should study and share experiences on the effects of integrating LA into everyday practice. They will have to implement better tools for non-(data mining) experts, which should optimally be embedded into their standard e-learning toolsets. LA tools should provide goal-oriented feedback to the different stakeholders for reflection, raising self-awareness, and decision support. For development of usable and useful LA tools it could be helpful to propose guidelines and design patterns. Also, faculty and student involvement will be the key to a wider user acceptance, which is required if LA tools are to serve the intended objective of improving teaching and learning. The number and hierarchy of stakeholders bears conflicts. Data and conclusions might also be used in other than the intended ways. 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 TEL scenarios. Hence, integration of LA into everyday practice of the different stakeholders is a future challenge and should be accompanied by appropriate didactical and organizational frameworks.

And last but not least ethics, data privacy and stewardship issues need to be resolved. We need to detect and prevent data misuse, determine boundaries of analytics, preserve confidential user information, protect the identities of the users at all times, and think about our obligation to act on the basis of newly gained knowledge.

4.3. 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, and reflection.

4.3.1. 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.

4.3.2. 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 accomplishments. 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.

4.3.3. Tutoring and Mentoring: Tutoring is mainly concerned with helping students with their learning (assignments), 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.

4.3.4. Assessment and feedback: The objective is to support the (self-)assessment of improved efficiency and effectiveness of the learning process. Important is also 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.

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

4.3.6. 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. 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 examine 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.

4.3.7. Reflection: Analytics can be a valuable tool to 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).

The above mentioned various objectives in LA are sometimes difficult to measure and need a tailored set of performance indicators and metrics. Moreover, we need to define new metrics beyond grades and graduation rates in order to support different types of learning including self-organized learning, network learning, informal learning, professional learning, and lifelong learning. The challenge is thus to define the right Objective / Indicator / Metric (OIM) triple before starting the LA exercise.

Furthermore, to achieve particular learner-centered LA objectives such as intelligent feedback, adaptation, personalization, or recommendation, learner profiling is a crucial task. The challenge is to create a thorough “polycontextual” learner profile that can be used to trigger effective intervention, personalization, or recommendation actions. This is a highly challenging task since learner activities are often distributed over open and increasingly complex learning environments and legal issues related to data privacy have to be taken into account. The capacity to build a detailed picture of the learner activities across a broader learning context beyond the LMS would provide a more accurate analytics results. A big challenge to tackle here is lifelong learner modeling (Kay and Kummerfeld, 2011). The aim is that data gathered from different (learning) environments would be fed into a personal lifelong learner model, which would be the “store” where the learner can archive all learning activities throughout her life. Thus, this model must store very heterogeneous data. It would also contain both long-term and short-term goals of the learner.

4.4. Methods (How?)

LA applies different 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).

4.4.1. 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, percentage of material read. These statistical tools often generate simple statistical operations such as average, mean, and standard deviation.

4.4.2. 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 with dashboards that graphically show different performance indicators.

4.4.3. 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). 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).

Classification is the process of finding (learning) a function (or model) that describes and distinguishes data classes or concepts, for the purpose of being able to use the function to predict the class of objects whose class label is unknown. The function is also called a classification model, a predictive model or simply a classifier. Popular classification methods include decision trees, neural networks, naive Bayesian classification, support vector machines (SVM), and k-nearest neighbor classification (KNN). While classification predicts categorical (discrete) labels, prediction models continuous-valued functions; i.e. it is used to predict missing numerical data values rather than class labels. Regression analysis is a statistical methodology often used for numeric prediction. Classification is also called supervised learning because the data objects used for learning (called the training data) are labeled with pre-defined classes.

Clustering (unsupervised learning) contrasts with classification (supervised learning) in that the class label of each training object is not known in advance. Clustering is the process of organizing the data objects into groups or clusters, so that objects within a cluster are "similar" to each other and "dissimilar" to objects in other clusters. Similarity is commonly defined in terms of how close the objects are in space, based on a distance function. In general, the major clustering methods can be classified into the following categories: partitioning methods, hierarchical methods, and density-based methods.

Partitioning methods construct partitions (or groups) of the data, where each partition represents a cluster. A partitioning method starts with an initial random partitioning. It then uses an iterative relocation technique that attempts to improve the partitioning by moving objects from one group to another. Partitioning methods use popular heuristic methods, such as the k-means algorithm (each cluster is represented by the mean value of the objects in the cluster) and the k-medoids algorithm (each cluster is represented by one of the objects in the cluster).

Hierarchical methods create a hierarchical decomposition (a tree of clusters) of the given set of data objects. A hierarchical method can be either agglomerative or divisive, depending on whether the hierarchical decomposition is formed in a bottom-up (merging) or top-down (splitting) manner.

Density-based methods view clusters as dense regions of objects in the data space that are separated by regions of low density (representing noise). A density-based method continues growing a given cluster as long as the density (number of objects) in the neighborhood exceeds some threshold; i.e. for each object within a given cluster, the neighborhood of a given radius has to contain at least a minimum number of objects. Such a method can be used to filter out noise (outliers) and discover clusters of arbitrary shape. DBSCAN and its extension OPTICS are typical density-based methods.

Association rule mining leads to the discovery of interesting associations and correlations within data. A typical example of association rule mining is the market basket analysis, which aims to discover how items purchased by customers in a store are associated. Thereby, patterns are represented in the form of association rules (e.g., computer -> antivirus_software [support = 10%, confidence = 80%]). Rule support and confidence are two measures of rule interestingness. The rule above, for instance, says that 10% customers buy computer and antivirus_software together, and those who buy computer also buy antivirus_software 80% of the time. Popular methods for mining association rules are the Apriori algorithm and frequent-pattern trees (FP-tree).

4.4.4. 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).

In sum, different techniques can be used depending on the objectives of the analytics task. The challenge is to design and develop usable and useful statistical, visualization, filtering, and mining tools which 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. Romero and Ventura (2010), for instance, point out that these tools can use a default algorithm for each mining task and parameter-free algorithms to simplify the configuration and execution of the mining tasks for different users. For balancing flexibility and usability of the tools, guidelines on good interaction design should be taken into account.

Currently many of the systems are data rich, but information poor. Researchers need to find pedagogically useful indicators, predictions and recommendations by evaluating the quality of analytics results in practice. How can available raw data be converted into actionable knowledge to the user? How can it become a trustworthy empirical base for decision-making?

A future challenge is also to develop effective LA tools that can be integrated within the learning environments and systems. Effective analytics tools are thus those, which minimize the time frame between analysis and action. Ideally, tools can return answers within seconds to allow for an exploratory, real-time user experience, and to enable data exploration and visualization manipulation based on individual research interests of stakeholders.

Furthermore, performance, scalability, and extensibility should be taken into account, i.e. tools should allow for incremental extension of data volume and analytics functionality after a system has been deployed.

Appropriate visualizations could make a significant contribution to understanding the large amounts of educational data. The goal of information visualization is to transform data into knowledge. This should allow gaining insights into processes and relationships of teaching and learning, which are the basis for activities of improvement.

Another challenge is the support of mixed method approaches. The LA process basically starts with asking a research question and proceeds with choosing appropriate methods. Many researchers and educators of the TEL domain use mixed method approaches to answer their research questions (Dyckhoff, 2010). Not only do they integrate different quantitative methods to increase the robustness of research but also qualitative methods, like interview techniques or focus groups, are sometimes used to answer the question of "why" something was observed. While quantitative methods and data mining techniques may show trends, correlations, connections, clusters or structures occurring in the data, qualitative methods might provide additional information on reasons. Therefore they can support the interpretation of analytic results and should be integrated. As today's learning environments are becoming more complex mixed methods are becoming essential for LA purposes.

5. A Review of the State-of-the-Art

The number of publications around LA and its related areas has grown rapidly in the last few years. In this section, we review the most relevant studies in this field in 2010 and 2011 and map them to the four dimensions of the LA reference model discussed in section 4. In this paper, we mainly reviewed studies from the LAK11, EDM10, EDM11 conferences (Long and Siemens, 2011; Baker et al., 2010; Pechenizkiy et al. 2011), which discussed concrete LA applications (18 in LAK11, 22 in EDM10, 20 in EDM11). Theoretical and vision papers are not covered in this review. To note that LA publications in 2010 and 2011 are not limited solely to the aforementioned ones. Several other conferences (e.g. AIED, ITS, UM, UMAP, ECTEL, RecSys) have included special tracks addressing LA and its related topics. These specialized tracks are, however, converging toward more general conferences (e.g. LAK and EDM) bringing together researchers from different disciplines. We therefore restricted our literature review to these two conferences.

5.1. Data and Environments (What?)

The LA tools that have been proposed in the reviewed literature use different data sources. We classified the data environments into seven main categories: student information systems (SIS), social media (e.g. chat, networking, conferencing), web-based courses, traditional learning management systems (LMS), adaptive intelligent educational systems (including intelligent tutoring systems (ITS) and adaptive hypermedia systems (AHS), personal learning environments (PLE), and open data sets. As illustrated in Figure 3, the tools are often linked to a centralized web-based learning systems; i.e. Adaptive system/ITS, web-based course, or LMS (85% of papers). Only two studies use open data sets. Baker and Gowda (2010) use data publicly available from the PSLC DataShop and Verbert et al. (2011) explore different data sets that capture learner interactions with TEL tools and resources (e.g. Mendeley, APOSDLE DS, ReMashed, MACE). All other studies use their own educational data collected in log files. By contrast, the review of the EDM literature in 2008 and 2009 revealed that 14% of the papers used publicly available data (Baker and Yacef, 2009). In our review, the decrease in usage of open data sets is likely a reflection of the privacy issues that arise with each LA application. To note also that the number of LA applications that use data from a student information system has significantly decreased as compared to early academic analytics studies which focused more on improving student retention and graduation rates. Furthermore, there is no study, which uses a PLE as a data source for LA. This can be explained by the fact that PLE is a relatively new research topic as compared to LMS or adaptive systems. We expect, however, that future LA applications will increasingly capture and mine data collected in PLEs, as a result of a shift in focus in the last few years from centralized learning systems to open learning environments (e.g., massive open online courses, MOOC) and from adaptation to personalization. Capturing data in open learning environments would require a multimodal net of wearable sensors, environmental sensors (e.g. to capture presence, sound etc.) and software sensors (desktop-based, Web-based or mobile-based).

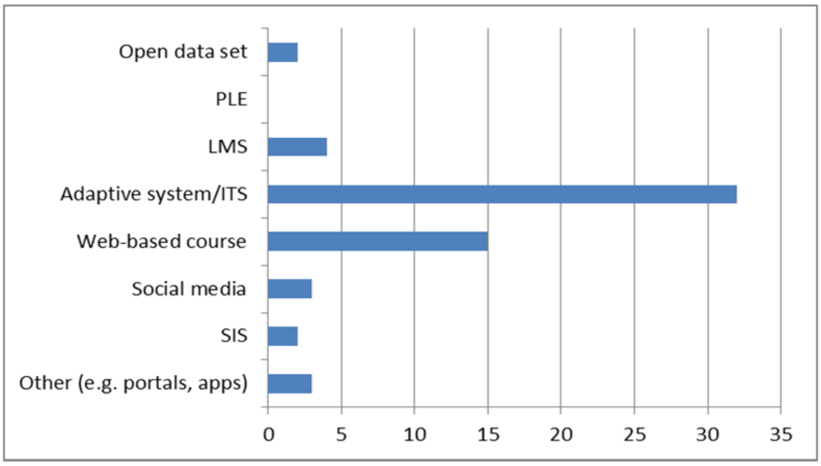

Figure 3. Data and Environments

5.2. Stakeholders (Who?)

Figure 4 shows the reviewed LA studies grouped by stakeholders. Most of the studies target intelligent tutors in adaptive systems (48%) or researchers/system designers (30%). Only few studies involve the teacher by supporting the monitoring of students' activities (e.g. Bakharia and Dawson; 2010; Romero et al., 2010b) or the learner by providing feedback (e.g. Clow and Makriyanni, 2010; Dominguez et al., 2010) or generating recommendations (Niemann et al., 2011). Future LA studies will need to take into account the pedagogical issues and as a consequence involve the learner and the teacher as key stakeholders. We believe that future studies will put a heavier emphasis on learners, as there is currently an increasing interest in self-directed learning and personal learning environments. In self-directed learning scenarios, learners should be at the center of the analytics process, in which they are shown their position and actions within the learning environment and consciously choose actions and set goals to achieve their learning objectives.

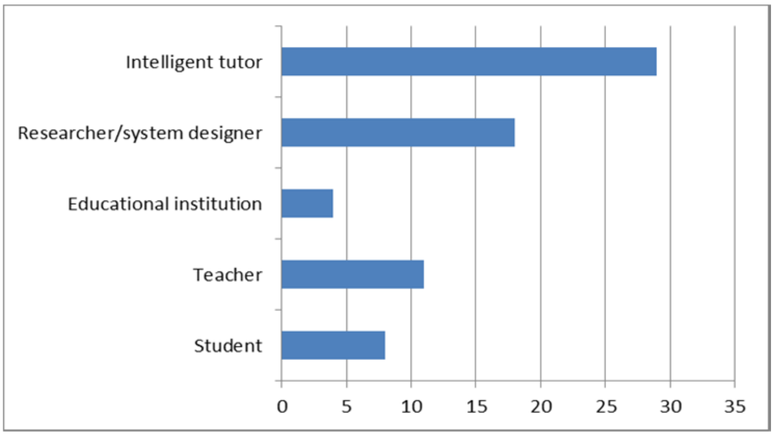

Figure 4. Stakeholders

5.3. Methods (How?)

As discussed above, we can categorize LA methods into the following classes: statistics, information visualization (IV), data mining (DM), and social network analysis (SNA). Figure 5 shows the reviewed literature classified according to the LA techniques used. Note that some studies applied different methods and can thus be found in multiple categories.

In their review of the EDM literature between 1995 and 2005, Romero and Ventura (2007) conclude that association rule mining and sequential pattern mining were the most prominent methods (43% of papers), followed by classification and prediction methods (28 % of papers) (as cited in Baker and Yacef, 2009). This pattern changed in EDM research in 2008 and 2009. In their review of the papers from EDM08 (Baker et al., 2008) and EDM09 (Barnes et al., 2009), Baker and Yacef (2009) note that classification and prediction methods moved to the dominant position (42% of papers), while only 9% of papers involved association rule mining methods.

This pattern persists in our review. The most used LA technique remains classification and prediction (38% of papers). This can be explained by the higher number of EDM publications included in the review. If we restrict our focus to LAK11 publications, we see a very different pattern. Only two studies applied data mining techniques (i.e. K-means clustering and Bayesian networks) (Brooks et al., 2011; Fanscali, 2011). All other studies focused on visualizations (50%), followed by statistics (39%) and SNA (22%). We noted that, whereas EDM research is dominated by the application of data mining techniques in adaptive educational systems, LAK research puts a heavier emphasis on the practical aspects of the LA exercise and addresses new forms of learning (e.g. open and network learning), which explains the rise in usage of new methods, such as SNA.

A significant number of the reviewed literatures (25% of papers) use statistics tools. Statistics are often augmented with visualizations techniques attempting to present the statistical results in a useful way.

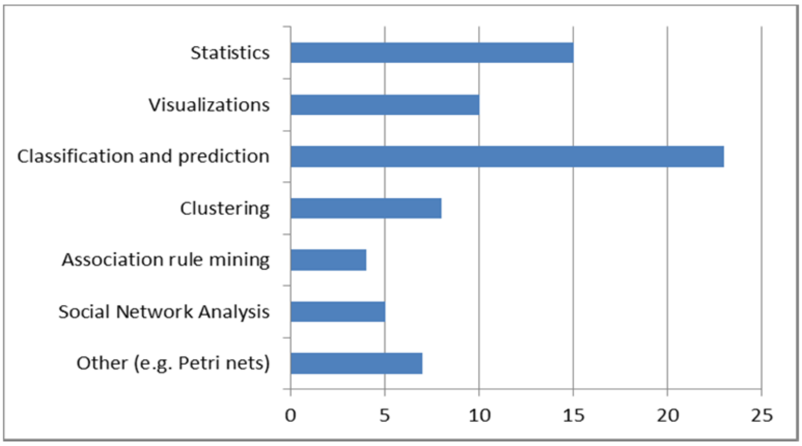

Figure 5. Methods

5.4. Objectives (Why?)

Figure 6 shows the number of publications grouped by objectives. In the next sections, we investigate the LA objectives in the reviewed literature along with the techniques applied to support these objectives.

The most applied objective in the reviewed literature is adaptation (40% of papers), as a result of a focus on adaptive and intelligent web-based educational systems in the last years. A common task in these studies is student modeling. Different data mining techniques have been used to build a student model. Classification and prediction algorithms (Bayesian networks, Expectation Maximization, hidden Markov models, decision trees, performance factor analysis) have been developed to predict the knowledge levels, behaviors, goals and mental states of students (cf. Fincham et al., 2010; Gong and Beck, 2011; Nooraei et al., 2011). Clustering methods (e.g. k-means) have also been proposed to support the student modeling task (Kardan and Conati, 2011; Nugent et al., 2010).

Another key trend in LA tasks is monitoring and analysis which is the second most applied objective in our review, with 33% of papers. This trend is more visible in LAK research, with 78% of LAK papers involving a monitoring and/or analysis component in their LA solutions. Statistics and visualizations are the main techniques that have been used to achieve this objective. Different statistics tools and visualization techniques have been applied, e.g., to monitor the activity on a site and the usefulness of a reputation system (Clow and Makriyanni, 2011), to inspect students' behavior and learning in project-based, unscripted constructionist learning environments (Blikstein, 2011), or to analyze of portal usage in relation to specific training events/days (Palavitsinis et al., 2011). Different data mining techniques have also been applied for monitoring and analysis. Jeong et al. (2010), for instance, use the hidden Markov model approach for exploratory sequence analysis to demonstrate that the high-performing students have more linear learning behaviors, and that their behaviors remain consistent across different study modules. And, Romero et al. (2010b) explore the application of association rule mining to student data stored in a large Moodle repository to monitor and analyze information about infrequent student behavior.

Assessment and feedback are less common objectives in the reviewed literature (13% of papers). Several methods have been applied for assessment and feedback but the most common are statistics and visualizations. For instance, Clow and Makriyannis (2011) use scientific scores in a reputation system to provide feedback on the scientific expertise of learners in an informal learning context. Vatrapu et al. (2011) propose a triadic model of teaching analytics, consisting of a teacher, a visual analytics expert, and a design-based research expert to support teachers in assessing students’ learning activities by using a range of visual analytics tools. Feng and Heffernan (2010) develop different metrics for dynamic assessment that measures student accuracy, speed, attempts, and help-seeking behaviors in a tutoring system. Data mining techniques have also been used for assessment and feedback. Dominguez et al. (2010) use clustering and association rule mining to provide automatic feedback to students who are completing programming exercises. Xiong et al. (2010) propose a classification approach based on decision trees to generate automatic assessment of students' reviewing performance in peer-review systems.

Though not prominent as it was the case in early academic analytics studies, 12% of papers in our review focus on the prediction of students' performance as a prior step to intervention. Different data mining techniques have been applied for predicting students' performance. Fancsali (2011), e.g., uses Bayesian networks to determine the predictors and causes of student learning outcomes given the records of their activities and interactions. Qui et al. (2011) consider the effect of time and show that a significant improvement of Bayesian Knowledge Tracing can be obtained on prediction by also modeling forgetting. Furthermore, Thai-Nghe et al. (2011), for instance, propose to use tensor factorization for forecasting student performance to personalize the prediction for each student given the task.

In general, we notice that in current LA practices the focus remain on adapting content and hardly the mirroring of the learning actions to teachers/mentors and learners, for effective feedback, mentoring, and reflection. Future LA practices need to follow objectives that put the learners and the teachers/mentors at the center of the analytics tasks by proposing solutions that (1) empower learners to reflect and act upon feedback about their learning actions and (2) keep teachers/mentors in the feedback loop to let them intervene on the learning process immediately end effectively.

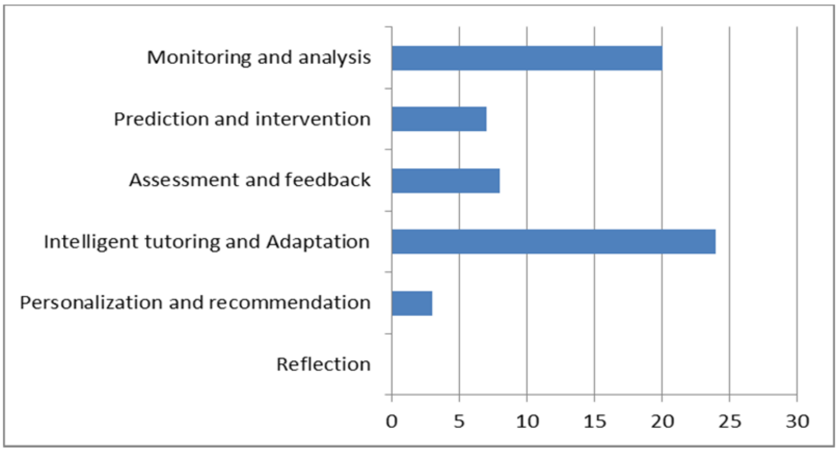

Figure 6. Objectives

6. Conclusion and Future Work

Learning analytics (LA) has attracted a great deal of attention in TEL in recent years as educational institutions and researchers are increasingly seeing the potential that LA has to support the learning process. LA approaches share a movement from data to analysis to action to learning. LA is an interdisciplinary field in which several related research areas converge. In this paper, we addressed the relationship between LA and its related fields. We also discussed a reference model that provides a classification schema of LA solutions based on four dimensions: data and environments (what?), stakeholders (who?), objectives (why?), and methods (how?). Based upon this classification, we reviewed recent literature related to this field. The review showed that (1) centralized web-based learning systems (e.g. ITS, LMS) represent the most widely used data source for LA, (2) most of the current LA applications are oriented toward intelligent tutors or researchers/system designers, (3) the most commonly applied objectives are adaptation and monitoring/analysis, and (4) the most frequently used LA techniques are classification and prediction. We believe that in the future these patterns will change, as the focus of LA will shift toward more open, networked, personalized and lifelong learning environments. LA further requires key stakeholders to address a number of challenges, including questions about handling increasing data volume, heterogeneity, fragmentation, system interoperability, integration, performance, scalability, extensibility, real-time operation, reliability, usability, finding meaningful indicators/metrics and appropriate information visualization, supporting mixed-method approaches (quantitative and qualitative), data privacy, stewardship, ethics, and integration of LA into everyday practice. These challenges will need to be addressed as the understanding of the technical and pedagogical issues surrounding LA evolves.

References

Adomavicius, G. & Tuzhilin, A. (2005). Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Transactions on Knowledge and Data Engineering*, 17(6), 734-749.

Altrichter, H., Posch, P. & Somekh, B. (1996). *Teachers Investigate their Work: An Introduction to the Methods of Action Research*, Routledge, London.

Arnold, K., 2010. Signals: Applying Academic Analytics. EDUCAUSE Quarterly 33 (1), 10.

Baker, R.S.J.d., Barnes, T. & Beck, J.E. (2008). *Proceedings of the 1st International Conference on Educational Data Mining*.

Baker, R. S.J.d. & Yacef, K. (2009). The State of Educational Data Mining in 2009: A Review and Future Visions. *Journal of Educational Data Mining*, 1(1), 3-17.

Baker, R. S.J.d., Merceron, A. & Pavlik, P.I. (2010). *Proceedings of the 3rd International Conference on Educational Data Mining*. Retrieved February 10, 2012, from http://educationaldatamining.org/EDM2010/

Baker, R.S.J.d. & Gowda, S.M. (2010). *An Analysis of the Differences in the Frequency of Students' Disengagement in Urban, Rural, and Suburban High Schools*. In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 11-20). Pittsburgh, PA, USA.

Bakharia, A. & Dawson, S. (2011). *SNAPP: A Bird's-Eye View of Temporal Participant Interactions*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge. (pp. 168-173). NY, USA: ACM New York.

Barnes, T., Desmarais, M., Romero, C., Ventura, S. (2009). *Proceedings of the 2nd International Conference on Educational Data Mining*.

Blikstein, P. (2011). *Using learning analytics to assess students' behavior in open-ended programming tasks*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge (pp. 110-116). NY, USA: ACM New York.

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

Burgos, D. (2007). Integration of Adaptive Learning Processes with IMS Learning Design Considering Corporate Requirements. Project Deliverable, Deliverable 1.11 of the Prolearn EU/FP6 Network of Excellence, IST 507310.

Brooks, C., Epp, C.D., Logan, G. & Greer, J. (2011). *The Who, What, When, Where, and Why of Lecture Capture*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge. (pp. 86-92). NY, USA: ACM New York.

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

Campbell, J.P. & Oblinger, D.G. (2007). *Academic Analytics*. EDUCAUSE White Paper. Retrieved February 10, 2012 from http://www.educause.edu/ir/library/pdf/PUB6101.pdf

Chatti, M.A. (2010). Personalization in Technology Enhanced Learning: A Social Software Perspective. Shaker Verlag, November 2010. Dissertation, RWTH Aachen University.

Clow, D. & Makriyanni, E. (2011). *iSpot Analysed: Participatory Learning and Reputation*. In: Proceedings of the 1st International Conference Learning Analytics & Knowledge. Banff, Alberta, Canada (in Press).

Dominguez, A.K., Yacef, K. & Curran, J.R. (2010). *Data Mining for Generating Hints in a Python Tutor*. In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 91-100). Pittsburgh, PA, USA.

Dyckhoff, A.L. (2010). Towards Tools for Educational Action Research: A Meta-Analysis on Applied Research Questions. In: *Proceedings of the IADIS International Conference E-Learning 2010* (pp. 235-242). Freiburg, Germany.

EDUCAUSE (2010). 7 Things you should know about analytics. Retrieved February 10, 2012, from http://www.educause.edu/ir/library/pdf/ELI7059.pdf

Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential. Retrieved February 10, 2012, from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Fancsali, S. (2011). *Variable Construction for Predictive and Causal Modeling of Online Education Data*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge (pp. 54-63). NY, USA: ACM New York.

Feng, M. & Heffernan, N. (2010). *Can We Get Better Assessment From A Tutoring System Compared to Traditional Paper Testing? Can We Have Our Cake (Better Assessment) and Eat It too (Student Learning During the Test)?* In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 41-50). Pittsburgh, PA, USA.

Fincham, J.M., Anderson, J.R., Betts, S. & Ferris, J. (2010). *Using Neural Imaging and Cognitive Modeling to Infer Mental States while Using an Intelligent Tutoring System*. In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 51-60). Pittsburgh, PA, USA.

Goldstein, P.J. & Katz, R.N. (2005). *Academic Analytics: The Uses of Management Information and Technology in Higher Education*. ECAR Research Study, Vol. 8. Retrieved February 10, 2012, from http://www.educause.edu/ECAR/AcademicAnalyticsTheUsesofMana/158588

Gong, Y. & Beck, J. (2011). *Items, Skills, and Transfer Models: Which really matters for student modeling?* In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero , C., and Stamper, J. (Eds.) Proceedings of the 4th International Conference on Educational Data Mining (pp. 81-90). Eindhoven, the Netherlands.

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

Hinchey, P. (2008). *Action Research Primer*. Peter Lang, New York.

Jeong, H., Biswas, G., Johnson, J. & Howard, L. (2010). *Analysis of Productive Learning Behaviors in a Structured Inquiry Cycle Using Hidden Markov Models.* In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 81-90). Pittsburgh, PA, USA.

Johnson, L., Smith, R., Willis, H., Levine, A. & Haywood, K. (2011). *The 2011 Horizon Report*. Austin, Texas: The New Media Consortium.

Kardan, S. & Conati, C. (2011). *A Framework for Capturing Distinguishing User Interaction Behaviors in Novel Interfaces*. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero , C., and Stamper, J. (Eds.) Proceedings of the 4th International Conference on Educational Data Mining (pp. 159-168). Eindhoven, the Netherlands.

Kay, J. & Kummerfeld, B. (2011). Lifelong Learner Modeling. In P. J. Durlach and A. Lesgold (Eds.), Adaptive Technologies for Training and Education. Cambridge University Press.

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

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

McNiff, J. & Whitehead, J. (2006). *All you need to know about Action Research*. SAGE Publications Ltd., London.

Niemann, K., Schmitz, H.-C., Scheffel, M. & Wolpers, M. (2011). *Usage Contexts for Object Similarity: Exploratory Investigations*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge (pp. 81-85). NY, USA: ACM New York.

Nooraei, B., Pardos, Z.A., Heffernan, N.T. & Baker, R.S.J.d. (2011). *Less is More: Improving the Speed and Prediction Power of Knowledge Tracing by Using Less Data*. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero , C., and Stamper, J. (Eds.) Proceedings of the 4th International Conference on Educational Data Mining (pp. 101-110). Eindhoven, the Netherlands.

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

Norris, D., Baer, L., Leonard, J., Pugliese, L. & Lefrere, P. (2008b). *Framing Action Analytics and Putting Them to Work*. EDUCAUSE Review 43(1).

Nugent, R., Dean, N. & Ayers, E. (2010). *Skill Set Profile Clustering: The Empty K-Means Algorithm with Automatic Specification of Starting Cluster Centers*. In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr.

(Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 151-160). Pittsburgh, PA, USA.

Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero , C. & Stamper, J. (2011). *Proceedings of the 4th International Conference on Educational Data Mining*. Retrieved February 10, 2012, from http://educationaldatamining.org/EDM2011/

Palavitsinis, N., Protonotarios, V. & Manouselis, N. (2011). *Applying Analytics for a Learning Portal: The Organic. Edunet Case Study*. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge. (pp. 140-146). NY, USA: ACM New York.

Qiu, Y., Qi, Y, Lu, H. Pardos, Z. & Heffernan, N. (2011). *Does Time Matter? Modeling the Effect of Time with Bayesian Knowledge Tracing*. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero , C., and Stamper, J. (Eds.) Proceedings of the 4th International Conference on Educational Data Mining (pp. 139-148). Eindhoven, the Netherlands.

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

Romero, C. & Ventura, S. (2010). *Educational Data Mining: A Review of the State-of-the-Art*. IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 40(6), 601-618.

Romero, C., Ventura, S., Pechenizkiy, M. & Baker, R.S.J.d. (2010a). *Handbook of Educational Data Mining*, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series.

Romero, C., Romero, J.R., Luna, J.M. & Ventura, S. (2010b). *Mining Rare Association Rules from e-Learning Data*. In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 171-180). Pittsburgh, PA, USA.

Siemens, G. (2010). What are Learning Analytics? Retrieved February 10, 2012, from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Long, P. & Siemens, G. (2011). *Proceedings of the 1st International Conference on Learning Analytics & Knowledge*. NY, USA: ACM New York.

Suthers, D. & Rosen, D. (2011). *A unified framework for multi-level analysis of distributed learning*. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge. (pp. 64-74). NY, USA: ACM New York.

Tanes, Z., Arnold, K. E., King, A. S., Remnet, M. A. (2011). Using Signals for appropriate feedback: Perceptions and practices. Computers and Education 57 (4), pp. 2414–2422.

Thai-Nghe, N., Horváth, T. & Schmidt-Thieme, L. (2011). *Factorization Models for Forecasting Student Performance*. In: Pechenizkiy, M., Calders, T., Conati, C., Ventura, S., Romero , C., and Stamper, J. (Eds.) Proceedings of the 4th International Conference on Educational Data Mining (pp. 11-20). Eindhoven, the Netherlands.

Vatrapu, R., Teplovs, C. & Fujita, N. (2011). *Towards Visual Analytics for Teachers' Dynamic Diagnostic Pedagogical Decision-making*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge (pp. 93-98). NY, USA: ACM New York.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R. & Duval, E. (2011). *Dataset-driven Research for Improving Recommender Systems for Learning*. In: Proceedings of the 1st International Conference on Learning Analytics & Knowledge. Banff, Alberta, Canada (in Press).

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

Xiong, W., Litman, D. & Schunn, C. (2010). *Assessing Reviewer's Performance Based on Mining Problem Localization in Peer-Review Data*. In: Baker, R.S.J.d., Merceron, A., Pavlik, P.I. Jr. (Eds.) Proceedings of the 3rd International Conference on Educational Data Mining (pp. 211-220). Pittsburgh, PA, USA.