# Deliverable D3.1
# Review article about
# LA and EDM approaches

| | |
|---|---|
| Document Version: | V1.6 – October 31, 2014 |
| Document Status: | Final |
| Document Type: | Report |
| Diss. Level: | Public |
| Lead Partner: | TUGraz |
| Delivery Date: | M8 (October 2014) |


| | |
|---|---|
| Authors: | Christina Steiner (TU Graz) |
| | Michael Kickmeier-Rust (TU Graz) |
| | Mustafa Ali Türker (SEBIT) |

# EXECUTIVE SUMMARY

The purpose of this document is to review the state of the art in the field of learning analytics (LA) and educational data mining (EDM), to highlight recent trends, and to build a solid starting point for the project. This systematic review of the analytics process, available methodologies, and existing challenges in this field of research serves as a basis for elaborating the LA framework for LEA's BOX and for designing and developing the related general and competence-based learning analytics services.

LA and EDM research methods of extracting meaningful information from learning related data, with the aim of understanding and optimising learning, learning environments, and instruction. An overview on key topics of LA and EDM is given, starting from their basic definition, the identification of key dimensions in this field and an explanation of related research topics, like business analytics and academic analytics. After that, the LA process with its individual stages is outlined and the different dimensions of LA are discussed in more detail. This includes the identification of the key stakeholders of LA and – related to these different target groups – a description of the key objectives and applications of LA.

For LA to be successful, an emphasis needs to be put on identifying, collecting, processing, and analysing educationally relevant data. A summary of the different categories and sources of data is given and types of indicators used for LA are presented. The analytics methodologies employed in a concrete LA project depend on the kind of data collected, as well as on the target stakeholder group(s) and their objectives. The most common methods that are used to extract meaningful patterns from educational data are presented in this document. An increasing number of tools exist that implement these methods and provide support in pre-processing, analysing, and visualising data. A systematic overview of different categories of tools and their main purposes and characteristics is given and examples of each type are provided. This is complemented by a summary of the most recent trends in LA technologies. Game-based learning and virtual worlds are emerging technologies that are acknowledged for its positive impact on learners. The application of LA in these research areas constitutes another trend in the field of LA and is summarised as an excursus section in this deliverable.

Although much progress has been achieved in LA in the last years, there are still a number of great challenges to be addressed in future research. The existing research and practice gap is probably the most pressing one, which is related to a set of more specific challenges, including data integration from different sources and the implementation of meaningful and intuitive tools for teachers and learners. In addition, there is an urgent need for convincing empirical evidence on the positive impact and added value of LA for learning and teaching, to foster acceptance and adoption of LA technologies in educational practice.

Finally, this document delineates the main ideas of the LA approach of LEA's BOX, which will take advantage from and advance existing LA approaches and address several challenges existing in this field of research. Grounding on well-founded psycho-pedagogical frameworks, LEA's BOX will provide a holistic framework to effectively assess, monitor, and promote knowledge and competence. Based on the review provided by this deliverable, conceptual research on the LA approach will continue and will be translated into the implementation of a set of analytics and data mining services and their integration in the LEA's BOX platform.

# TABLE OF CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| BI | Business Intelligence |
| EDM | Educational data mining |
| ICT | Information and communication technology |
| ITS | Intelligent Tutoring System |
| LA | Learning analytics |
| LASG | Learning Analytics for and in Serious Games Workshop |
| LMS | Learning Management System |
| OECD | Organisation for Economic Co-operation and Development |
| SNA | Social Network Analysis |
| SNAPP | Social Networks Adapting Pedagogical Practice |

# PREFACE

Learning analytics, educational data mining, continuous formative e-assessment, all requires technology rich classrooms and requires collecting and systemizing information about activities and learning. The technological richness potentially offers a new and broad dimension of pedagogical possibilities (cf. Groff, 2013). A particular advantage is seen in conveying the 21st century skills and digital competencies. Thus, not surprisingly, some authors argue that the use of ICT in teaching and learning must be one of the key components of educational policies (Ferrari, Brecko, & Punie, 2014). And consequently, teacher training in ICT and digital skills, including e-skills, computer literacy, media literacy, and particularly analytic skills in assessing and reasoning over student data is a critical factor. Also, educators must have insight and willingness to apply ICT appropriately. This view is underpinned by the recent European educational policy (EC, 2013). Owston (2006) provides some evidence for such arguments and reports that teacher's digital skills have a significant impact on learning outcomes. Likewise, Krumsvik (2013, cited by Wasson & Hansen, in press) argued that digital competencies and formative e-assessment literacy are factors to ensure good learning outcomes.

Despite the highly acclaimed advantages of ICT and digital skills of teachers, European classroom reality is different, though. On the one hand we find a lack of technical infrastructure, which makes it difficult to collect the necessary data to conduct things like learning analytics, data mining, or formative e-assessment. On the other hand, we often find a rather negative attitude towards technology per se as well as the approach of collecting and analyzing data. These challenges are also reflected in the European survey of ICT use in schools (EC, 2013).

As a consequence, bearing the vision of learning analytics, educational data mining and formative e-assessment in mind, it is not only important to equip classrooms with technical infrastructure and teachers with general digital skills, to drive the vision and to gain from the opportunities we need to (i) communicate potential benefits to a broader audience (then mere academic communities, as it is often the case), (ii) we need to educate teachers in analytics and inquiry skills (as a special form of 21st century skill), (iii) we need to generate real pedagogical gain by not only providing technology but also didactical strategies, and finally (iv) we have to make the tools as simple, intuitive, and intelligent as possible and tailor them to concrete real-world demands.

The purpose of this document is to review the state of the art in the field of learning analytics, to highlight recent trends and build a solid starting point for the project.

# 1. INTRODUCTION

Work package 3 of LEA's BOX is concerned with research on learning analytics, based on the existing state of the art and on psycho-pedagogical frameworks (Competence-based Knowledge Space Theory, Formal Concept Analysis) serving as sound foundations for advancing and developing novel and competence-based learning analytics approaches. This desktop and conceptual research is translated into the implementation of a set of analytics and data mining services for the integration in the project's Web platform. This document provides a comprehensive overview of the state of the art in learning analytics (LA) and educational data mining (EDM) and results from the work accomplished in WP3 T3.1. The systematic review of the learning analytics process, existing methodologies, their benefits and drawbacks, and overall trend and challenges existing in the field serve as a theoretical basis for designing and developing the general and competence-based learning analytics services for LEA's BOX (T3.2 and T3.3).

A first short version of this review (Steiner, Kickmeier-Rust, & Albert, 2014) has been presented at the workshop "Learning Analytics for and in Serious Games (LASG)" at EC-TEL 2014, co-organised by the LEA's BOX project in cooperation with the GALA Network of Excellence (see Figure 1). The final version of this review will be submitted for journal publication.


**Figure 1:** The Learning Analytics for and in Serious Games Workshop at EC-TEL 2014 – a joint initiative of the GALA Network of Excellence and the LEA's BOX project.

## 1.1 Introductory Considerations

Assessment plays an important role in education; it is crucial to identify to what extent learning objectives have been met, and in order to be able to provide supportive or remedial interventions to learners and to inform and direct teaching. Assessment or evaluation, as it is also commonly denoted in an educational context, consists in the application of a range of methods for gathering and evaluating information about learning and instruction, with the purpose of making judgments of learners' work regarding courses or units of learning (Taras, 2005). Usually, a distinction between summative and formative assessment is made (Harlen & James, 1997; Shute, 2010). Summative assessment reflects the traditional approach of assessing educational outcomes and can be characterized as 'assessment of learning', i.e. measuring the achievement of learners in a systematic way (e.g. through standardised tests) and at certain intervals (e.g. at the end of the school year or marking period). Formative assessment can be described as 'assessment for learning'; it is an element and supporting tool of the teaching and learning process and has the positive intent of recognizing progress, promoting learning and planning next steps. Formative assessment may take different forms located on a continuum from formal (teacher-centric view, e.g. administering tasks to students) to informal (student-centric view, e.g. interactive classroom activities) (Shute, 2010).

The relevance of assessment, in particular formative assessment, as part of teaching becomes evident when considering the increasing numbers of dropouts in education. In Austria, for example, recently published statistical figures indicate that 154,000 adolescents of age 15 to 24 years are educational dropouts (John, 2014, July 15). In Europe dropouts are a hot topic, in general; for instance, it has been stated by Quinn (2013) that „too many students in the EU drop out before the end of their higher education course. This is a problem across the EU, as success in higher education is vital for jobs, social justice and economic growth" (p. 9). Based on current OECD numbers 17% of young people in Europe will not finish secondary education; only 40 % of European students complete university level education (OECD, 2013). Educational institutions and policy makers seek ways to reduce the number of early school-leavers who do not complete education programmes; reducing school drop-out rates and increasing completion rates of third level education constitute Europe 2020 targets (European Union, 2013). Summative and formative assessment constitute essential components of effective dropout prevention (Lee Goss & Andren, 2014); i.e. measuring existing knowledge and competencies, assessing and comparing performance, monitoring – and using that information as a basis for diagnoses and instructional support (Shute, 2010). As important as assessment itself is, of course, communicating the assessment result and using it for feedback to the learner (e.g. Shute, 2008). This is central especially for formative assessment; in fact, feedback can be considered an essential component and direct application of formative assessment (Shute, 2010).

The research areas of LA and EDM are part of a current trend in educational technology and aim at making sense of learning-related data. LA and EDM are considered to have great potential to support and advance educational assessment and feedback in educational practice, and to support establishing a deeper understanding of the learning and teaching process. They may also account for the different tools and technologies available for teaching and learning purposes today and the multifaceted data available from individual learners. LA and EDM can be described as educational application of big data (Johnson, Adams Becker, Estrada, & Freeman, 2014a) and consist in the research and implementation of methods on how to extract meaningful information from this big data. The main idea of learning analytics is not new, in essence, the aim is using as much information as possible about learners to understand the meaning of the data in terms of the learners' strengths, abilities, knowledge, weakness, learning progress, attitudes, and social networks, with the final goal of supporting learning and teaching, and providing the best and most appropriate personalized support.

LA and EDM are considered to have immense potential in fostering an evolution of education from a one-size-fits-all delivery approach to a flexible and responsive approach of instruction tailored to learner needs and interests (Johnson et al., 2013). They are seen as a key emerging technologies on education in the Horizon Reports for Higher Education and K-12 (Johnson, Adams, & Cummins, 2012; Johnson et al., 2014a, 2014b; Johnson, Adams Becker, Cummins, Estrada, & Freeman, 2013) in the last few years and have been described as a development that will dramatically shape the future of education (Long & Siemens, 2011).

A prerequisite for the adoption of learning analytics in teaching and learning practice is, of course, the accessibility and use of educational technologies allowing the collection of learning-related data. Today a multitude of different tools and system are available for educational purposes. With their application educational institutions have available and deal with increasingly large amounts of data, which can be used to extract additional value through LA (Ferguson, 2012). Although the use of information and communication technology (ICT) in schools is increasing in Europe, there are still some obstacles to its broad adoption, which are related to the general availability of the necessary technological infrastructure in schools, as well as to teachers attitudes towards using ICT in education (Wastiau, Blamire, Kearney, Quittre, Van de Gaer, & Monseur, 2013). Although teachers usually use mobile and digital technologies in their private live, they do so to a considerably smaller extent in their teaching practice (cf. D5.2 – Drnek, Türker, Steiner, Hillemann, & Kickmeier-Rust, 2014).

## 1.2 Structure of this document

This document gives an overview of LA and EDM, in general, and on existing methodologies in this field, in particular. Figure 2 illustrates the main topics of the deliverable. This review provides a basis for the further discussion, elaboration, and development of LA towards a holistic analytics approach that is shaped and implemented in the LEA's BOX project. The remainder of this deliverable is structured as follows: First, the basics on LA and EDM are presented (section 2), in terms of their key concepts, objectives, the general process of and data used in analytics, as well as the visualisation of results. Subsequently, in section 3 existing methodologies are presented – general methods used for analytics, and tools (for) implementing these methods. In addition to this, section 4 gives an overview of the latest trends with respect to LA technologies. An excursus to the use of LA in virtual worlds and serious games is presented as another trend and shows how LA is starting to develop also in these emerging educational technologies (section 5). This is followed by an overall discussion of drawbacks and challenges existing with current LA approaches (section 6), and reflections on the empirical evidence and application of LA in educational practice (section 7). Initial considerations on the LEA's box approach towards LA are outlined in section 8. Finally, a wrap up is given and next steps are discussed (section 9).

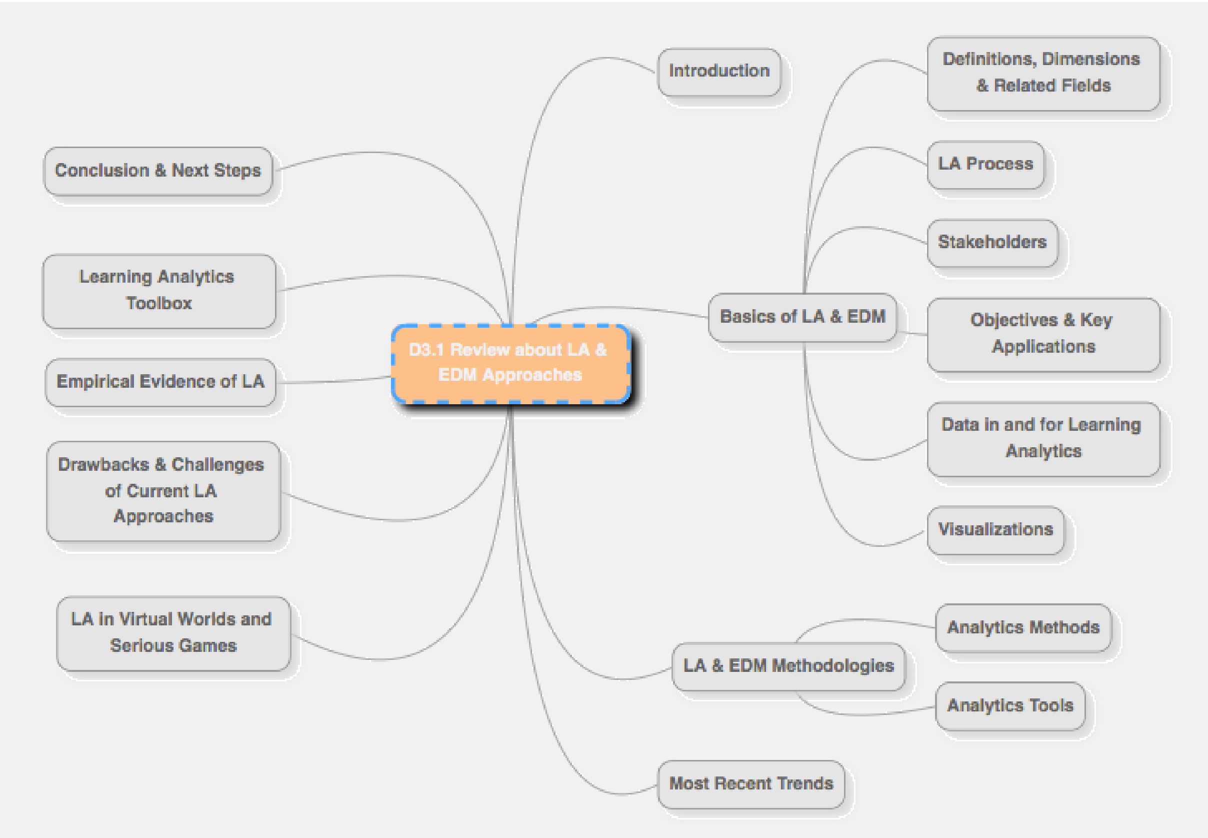


**Figure 2:** Overview of the contents of this deliverable.

# 2. BASICS OF LEARNING ANALYTICS AND EDUCATIONAL DATA MINING

This section gives an overview of the main relevant topics and areas of LA and EDM. After the presentation of how LA and EDM are defined and which other research concepts are related to them, we elaborate on the LA process and its individual steps. Subsequently, the main stakeholders are discussed, and the main objectives and applications of LA are presented. This is followed by a discussion of the learning data used, and of visualisations used in the context of LA.

## 2.1 Definitions, Dimensions, and Related Fields

### 2.1.1 Defining and Distinguishing LA and EDM

Deployment and use of learning or course management systems, web-based learning environments and learning tools produce a great variety of learning-related data and lead to educational institutions dealing with increasingly large amounts of data (Ferguson, 2012). In educational contexts, thus, a wide range of data about learners is potentially available today. A crucial question is how to make sense of these big data sets for assessment, learning, and teaching. Educational institutions so far have been commonly inefficient in making use of this data. In particular, the available data has traditionally been analysed with substantial delays, thus leading to delayed action and missed opportunities for interventions, like taking measures to reduce or avoid dropouts (Long & Siemens, 2011).

LA and EDM constitute related areas of research that aim at making sense of learning-related data; they deal with large data sets about learners and their contexts, in order to understand and develop learning (Ferguson, 2013). Both research areas are defined similarly and deal with similar concerns.

LA is defined by the Society of Learning Analytics Research[1] 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".**

Very similarly, the International Educational Data Mining Society[2] describes EDM as a

**"discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in".**

Apart from these, a range of other definitions exists. Although there might be slight differences, all definitions refer to the collection and analysis of data from learning processes, and share an emphasis on translating this data into meaningful actions to support and empower learning (Chatti, Dyckhoff, Schroeder, & Thüs, 2012).

In general, EDM and LA can be characterised as research areas with slightly different perspectives, but a significant overlap in their objectives and methods. The origins of EDM are usually dated back to the late 1990ies; LA emerged especially in the last decade (Ferguson, 2012; Romero & Ventura, 2010). While EDM has a focus on automated methods, in terms of automated analysis as well as applications in automated adaptation, LA also applies human-led methods or aims at leveraging human judgement to make sense of educational data and seeks applications in terms of using the derived information to empower and support learners and teachers (Baker & Siemens, in press;

[1] http://www.solaresearch.org/

[2] http://www.educationaldatamining.org/)

Romero & Ventura, 2013). In line with this, sometimes the two areas are described as having different roots, with the EDM community consisting mostly of researchers from the field of intelligent tutoring systems, and LA researchers having a greater focus on traditional learning systems and learning management systems (LMS). Interestingly, Chatti et al. (2012) describe EDM as focusing on typical data mining techniques, and LA as including also other methods like visualization tools and social network analysis; which is in contrast to Romero and Ventura (2010), who explicitly explains in his review that EDM includes typical data mining techniques, but also other approaches (like correlation, visualization etc.), which are not considered to be data mining in a strict sense. In any case, EDM is seen as rather focusing on the technical challenge of extracting information from learning-related data, and LA addressing more the educational challenge of optimizing learning (Ferguson, 2012). In general, LA can be seen as a more holistic approach (Baker & Siemens, in press), with the deployment of results from analytics and (cor-)responding action as important components in addition to pure analysis. John Behrens outlined in his keynote at the International Learning Analytics and Knowledge conference in 2012 (LAK2012) that EDM concentrates more on learning as a research topic, and LA has a more practical educational focus (Baker & Inventado, 2014). Eventually, both fields of research are closely related and share their interest in enhancing educational practice through researching data-intensive methodologies to education research. They are also both described conjointly in introductory and review articles, and sometimes terms seem to be even used interchangeably (Baker & Inventado, 2014; Romero & Ventura, 2010). In the remainder of this document the term LA will be used for referring to the wider research area and process of LA and EDM. We have given preference to the term LA, since in LEA's BOX the focus is on the educational application and relevance of the LA technologies developed in the project.

### 2.1.2 Key Dimensions

Figure 3 illustrates the main dimensions of learning analytics. Chatti et al. (2012) presented a reference model, in which they distinguished four main dimensions for LA, characterising the 'who', 'why', 'what', and 'how' of an LA project:

- **Who? – Stakeholders**: This refers to the people targeted by the analysis. This dimension is presented in section 2.3.
- **Why? – Objectives**: This dimension refers to the motivation for or goals of doing the analysis and is elaborated also in section 2.4.
- **What? - Data and environment**: This dimension refers to the kind of data that is gathered, managed, and used for analysis. This topic is discussed in more detail in section 2.5.
- **How? – Methods**: This dimension relates to the techniques and tools used for performing the analysis of collected data and is outlined in section 3.

In addition to these critical dimensions for the domain and application of LA, Greller and Drachlser (2012) identified two additional dimensions in their approach of defining a generic framework for LA:

- **External limitations**: This dimension refers to conventions (ethics, personal privacy, socially motivated limitations) and norms (legal and organisational constraints).
- **Internal limitations**: This dimension refers to relevant human factors, like competence (e.g. interpretation, critical thinking) and acceptance that may conflict with or complicate LA.

Concrete instantiations on these six dimensions of learning analytics characterize a specific application or use case of LA (Greller & Drachsler, 2012).

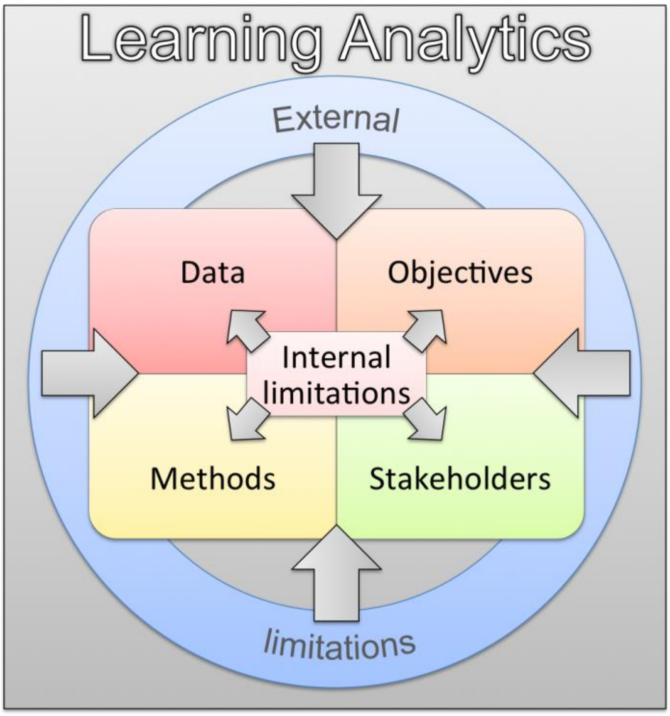


**Figure 3:** Dimensions of learning analytics.

**When characterising LA and LA applications in terms of the type of questions addressed, two other kinds kinds of dimensions are particularly relevant: time frame and innovation – 'Does analytics consider the past, present, or future?' and 'Does analytics lead to known information or produce new insight?' (Davenport, Harris, & Morison, 2010). From the combination of these two dimensions, different types of of questions result that may be addressed by LA – see**

Table 1. Questions related to information are about using information more effectively. Most current LA approaches focus on data about the past, reporting what has happened. Focusing on the past, however, means lost opportunities for intervention (Long & Siemens, 2011). Other analytics use past patterns and relate it to a predicted future; using forecasts and predictive modelling to identify indicators of success, failure or student drop out. A third approach is to take a more continuous and monitoring perspective, focusing on the present, for example generating alerts when an activity deviates from the normal pattern. Analytics aiming at gaining new insight are more challenging and usually require different and more sophisticated approaches. When focusing on the past, this might consist in trying to explain how things happened and why. Insight into the present related to the

derivation of recommendations of what to do now or next. Insight into the future results from the application of sophisticated methods to analyse the best/worst that can happen (Davenport et al., 2010). An approach towards information and insight about the present means to consider LA as part of the learning and teaching process, which is probably the most preferable way of using LA. In general, the aim of LA should be to go beyond answering questions related to information, especially to go beyond pure reporting about the past, but should strive producing additional insight on the learning and teaching process. An important question is also whether a LA approach is applicable in real-time or manually/on an on-demand basis (Van Harmelen & Workman, 2012), which is related to and also affects the potential application in terms of time-scale.

**Table 1:** Questions and points in time on which analytics may provide information and insight (adapted from Van Harmelen & Workman, 2012 and Davenport et al., 2010)

| | **Past** | **Present** | **Future** |
|---|---|---|---|
| **Information** | *What happened? (Reporting)* | *What is happening now? (Alerts)* | *What will happen? (Extrapolation)* |
| **Insight** | *How and why did it happen? (Modelling, experimental design)* | *What's the next best action? (Recommendation)* | *What's the best/worst that can happen? (Prediction, optimization, simulation)* |

LA can be done at different levels and characterised accordingly in macro-level, meso-level, and micro-level analytics (Buckingham Shum, 2012; MacNeill, Campbell, & Hawskey 2012, 2014). Macro-level analytics refers to cross-institutional analytics and the attempt to assess learning data on a longitudinal or lifelong learning level. Meso-level analytics are LA done at an institutional level. Macro- and meso-level analytics are actually what is also called 'academic analytics' (see section 2.1.3 below). LA at this level helps to establish an understanding of the success and potential challenges of an educational institution and may improve organizational productivity and decision-making. The term micro-level analytics characterises monitoring and interpretation of learning data of and for individual learners. This aims primarily to help learners understanding their own learning, to give recommendations for improvement and to identify risk for drop-out and potential interventions. Actually, these three layers are not independent from each other, but mutual dependencies, enrichment and integration is actually taking place, such that micro- and meso- level analytics may lead to transforming the educational system and pedagogical approaches on a macro-level. (Buckingham Shum, 2012).

### 2.1.3 Related concepts and research areas

LA is closely tied with and draws from a range of related fields of research; the most important are shortly outlined below.

Analytics has been spreading over the last years and decades in different domains. Researching methodologies on how to extract meaningful information from big data has a long tradition in natural sciences, and more recently has become an important part of business in terms of web and business analytics, and reached the field of learning science comparably late (Baker & Siemens, in press). In fact, LA is commonly described as having its roots in the commercial sector: *Web and business analytics* serve identifying consumer activities and preferences, analysing consumer trends etc., with the goal of tailoring product actions and advertising to consumers (Johnson et al., 2013). Business analytics are applications and technologies targeted to gather, store, analyse and provide access to data, with the aim of supporting better business decisions. The idea of LA is to leverage analytics methods to support decisions on learning and learning experiences. The main reasons for the growing interest and application of analytics approaches in educational domains are: Increasing amounts of data are available through the use of mobile, digital, and online technologies and their growing use for educational purposes. Information and experience on how to track this data is accessible, and standardized data formats are available. This, as well as the increasing computational power given today has nurtured work on analytics tools that support capturing, organizing, and filtering data, as well as tools that support the application of analytics and data mining methods (Baker & Siemens, in press). The idea is, in the end, to use learning data for recommendations (of learning resources, activities, people) and to adapt instruction in a similar manner as it is done with books, music, entertainment etc. in e-commerce (Johnson et al., 2014a). LA therefore has also strong relations to *recommender systems* (Adomavicius & Tuzhilin, 2005), *adaptive learning environments* and *intelligent tutoring systems* (ITS; Brusilovsky & Peylo, 2003), and the goals of these research areas. Apart from the idea of using LA for automated customisation and adaptation, feeding back LA results to learners and teachers to foster reflection on learning has also high relevance for *self-regulated learning* (Zimmerman, 2002) (see also section 2.2 below).

Further research fields in the educational context that are linked to LA and share similar objectives are academic analytics and action research. Academic analytics emphasizes the exploitation and analysis of educational data for educational institutions and authorities at regional, national or international, and governmental level. Academic analytics is less specific than learning analytics, since the focus is more on an analysis at institutional level instead of an analysis of the learning process itself (Long & Siemens, 2011). Academic analytics can help educational institutions in better fulfilling their educational mission (Campbell & Oblinger, 2007). Academic analytics and LA initially evolved conjointly; in recent years both areas are gradually developing as separate research areas (Ferguson, 2012), but overlaps between them naturally remain. Van Harmelen and Workman (2012) pinpoint learning analytics as aiming at optimising learning and teaching per se, while academic (or educational) analytics is focusing on optimising activities around learning and teaching, like recruitment or selection of students etc. Buckingham Shum (2011) describes academic analytics as being more inspired from well-established fields of business analysis and intelligence, while LA is more concentrated on the micro-level of learner interactions. Action research is generally described as reflective teaching practice, where instructors analyse, self-reflect, evaluate, and regulate their didactical methods and learning resources provided to students, with the aim of improving teaching practice and assuring quality (Chatti et al. 2012; Dyckhoff, Lukarov, Muslim, Chatti, & Schroeder, 2013). This is in line with the idea of teachers using LA for improving their teaching. The main purpose of action research is quality assurance and improvement of instruction; starting points are usually research questions arising from teaching practice (Chatti et al., 2012).

Aside from using analytics for gaining better understanding on how students learn, analytics can also be applied to educational content in terms of analysing how educational resources are used, by whom, and in what context (MacNeill et al., 2014). This approach is referred to as *educational content analysis*. The obtained results are also called ‘paradata’; it may complement metadata of content objects and is created while resources are being (re-)used, adapted, contextualised, tagged, shared etc.

## 2.2 The LA process

Baker (2007) used the so-called ‘knowledge continuum’ as a metaphor to describe the LA process. Raw data is transformed into information by giving the data meaning; this information – through analysis and synthesis – is used to answer questions and in this way becomes knowledge (or insight). Knowledge is applied through predictive analysis and development of actionable knowledge to establish and achieve goals and, in this way becomes wisdom.

The process of LA is usually described as a multi-step, cyclical process consisting of three main stages, as illustrated in Figure 4: data collection and pre-processing, analytics and action, and post-processing (Chatti et al., 2012). Data collection and pre-processing refers to the gathering of educational data from different learning systems and applications. Since the collected data may be very extensive or cover irrelevant information, the data is prepared and translated into an appropriate format for the next step. The analytics and action phase denotes the actual application of analytic methods, to extract meaningful patterns and information from the data. This step also includes visualization of the derived information and action, like prediction, assessment, adaptation, personalization, and reflection. Post processing refers to the idea of continually improving analytics, by refining analytics methods or using new methods, including new data sources etc.

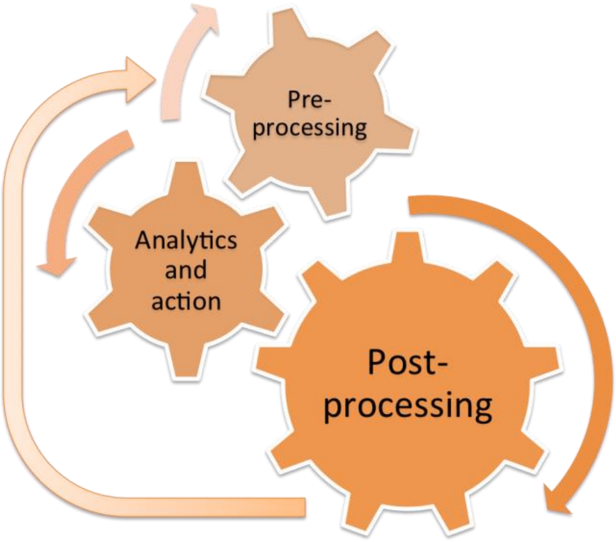


**Figure 4:** The basic LA process.

A more fine granular approach of describing the LA process is to distinguish five steps: capturing, reporting, predicting, acting, and refining (Campbell & Oblinger, 2007; Pardo, 2014). Comparing different models and frameworks describing the LA process, Elias (2011) additionally identified the stage of 'selecting' before and 'aggregating' after data capture. The individual steps are presented in more detail below. Figure 5 presents an illustration integrating the basic LA process (cf. Figure 2), with more fine granular descriptions (Campbell & Oblinger, 2007, Pardo, 2014, Elias, 2011) and the consideration of the knowledge continuum (Baker, 2007).

- **Select**: At the beginning of the LA process, it is necessary to define a goal and to carefully select the data to be collected and used in line with this goal (Elias, 2011). Important steps regarding data selection are: to determine what kind of data is available (and in which form), and what data provides useful insights. Relevant questions in this context – especially when aiming at continuous/real-time LA are the questions of storage, granularity, and retention (Campbell & Oblinger, 2007).
- **Capture**: This stage refers to collecting and storing learning-related data, potentially from different sources. Data capture and representation is actually influenced by the pedagogical strategy underpinning the educational scenario from which data is collected (Pardo, 2014). The major source of learning data used in LA today are logs of events occurring in an LMS; but actually obtaining data only from the LMS does not consider learning events occurring outside the LMS which is a limitation of current LA approaches. Triangulating data from multiple sources increased the quality of LA predictions and results (Pardo, 2014). In this way a richer picture of the learning process, progress, and performance can be obtained (Elias, 2011).
- **Aggregate**: Since the data collected may be noisy, may include data on irrelevant attributes or in different formats, or may come from different, heterogeneous sources, the aggregation

stage also includes the preparation and pre-processing of data, integration/centralisation and encoding in order to make the data ready for further processing. Developing a unified representation of learning data is actually a challenge in LA. The main goal is thereby to come up with a representation format that keeps a balance between being generic, while keeping expressive power (Pardo, 2014).

- **Report**: A minimal LA experience consists in simply capturing and reporting back information (Pardo, 2014). This step consists in processing the data in terms of visualisations or algorithms that summarise or combine data and reporting it back to stakeholders. This may be done continuously, in terms of real-time computation, or periodically. Reports largely rely on visual approaches to represent information. Dashboards are a popular type of visualisation used in LA; they provide a comprehensive view on a large amount of information (Pardo, 2014). With dashboards, though, it is critical to provide actually useful information to stakeholders (Elias, 2011).
- **Predict**: This refers to using the data to provide more sophisticated reports and answering previously defined questions. Concretely, this means that learning data captured about certain events is used to predict events in the future. Two large categories of prediction methods are statistical inference (e.g. linear regression) and machine learning (Pardo, 2014).
- **Act**: This stage consists in actions taken based on the previous phases. It includes automatic (e.g. adapting course material) as well as manually deployed (e.g. talking to a student) actions that change or affect any aspect of the learning activity (Pardo, 2014). This step of the process actually highlights that LA is not a purely technical solutions; rather a successful analytics approach and solution involves human-decision making and action as much as technical components (Van Harmelen & Workman, 2012). The number, type and intensity of interventions largely depends on the educational context and the LA project. In any case, an evaluation of the impact of interventions should be planned (Campbell & Oblinger, 2007).
- **Refine**: An analytics solution will usually not be optimal at first, such that its evaluation and consideration of technical and social aspects shall form the basis for on-going improvement (Van Harmelen & Workman, 2012). The final step of the LA process thus relates to the idea of continuously reviewing and adjusting the LA process and stages, in order to increase the accuracy of results and suitability of the LA process and maximise impact. This step is referred to more as a philosophical proposition and it is definitely the least documented in the literature (Pardo, 2014).

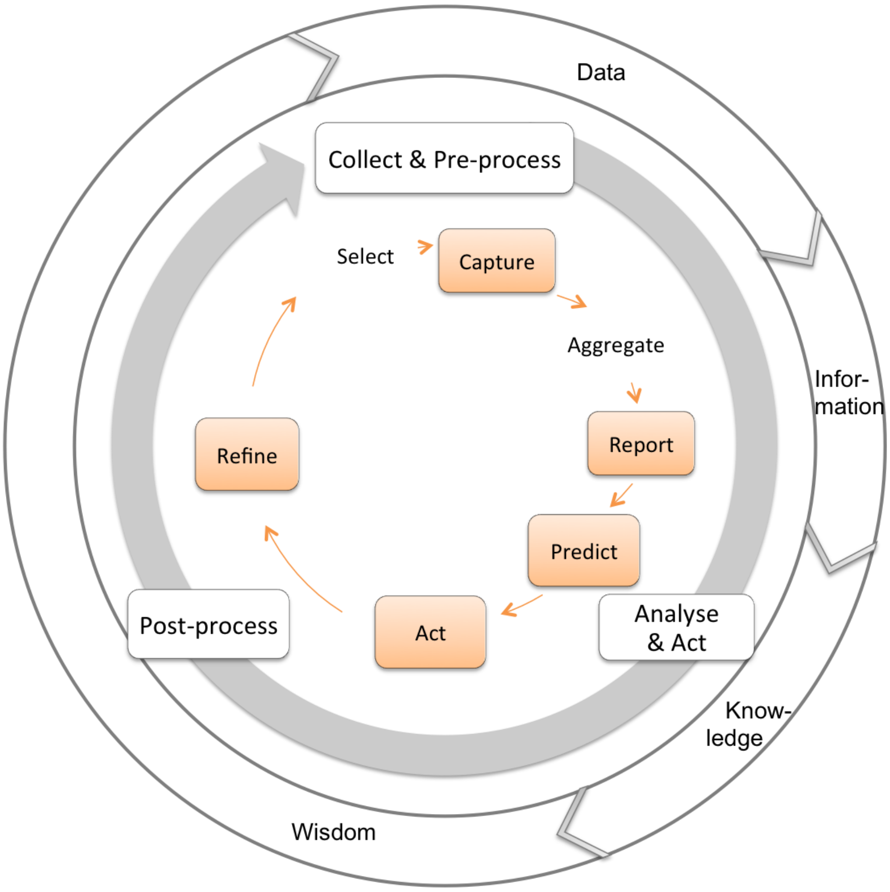


**Figure 5:** Integrated overview on different models describing the LA process.

## 2.3 STAKEHOLDERS

LA can be oriented towards and carried out for different stakeholders, who have different expectations, needs and goals towards the analytics process and its outcomes. The key stakeholders of LA are certainly teachers and learners; but there are actually more groups of people involved and interested in LA, with other objectives and perspectives on the data (Dyckhoff et al., 2013; Romero & Ventura, 2013). Those other groups of stakeholders are educational institutions and administrators, as well as course developers and training providers, but also researchers (Romero & Ventura, 2010).

Learners will mainly be interested in LA in how to become better learners, how analytics might help improving their performance, participation, or grades, to increase their awareness on the learning process and their own knowledge and competence. Teachers can use learning analytics to understand and improve the effectiveness of their teaching practice, to understand the usefulness of learning resources, activities and tools, to monitor the progress of their students, to identify problems and adapt their teaching offerings. Educational institutions and administrators will be interested in LA at an institutional level (note: in this case, strictly speaking, it is academic analytics that is carried out – see section 2.1.3 above) to support educational decision making and planning, to adapt curricula and organise educational offers. Analytics may serve identifying students at risk and selecting most qualified applicants. Training developers and providers are interested in LA to evaluate the effectiveness of the trainings provided and to understand how course material is used. This may assist further planning of training offers, the construction and reuse of learning material. Educational researchers will likely apply LA to investigate the effects and benefits of instructional strategies, educational tools and interventions. LA researchers are mainly interested in finding out about the impact that LA may have on learning and teaching, and in how to further refine LA methodologies.

It has been highlighted, that the LA dimension related to stakeholders includes not only data clients, i.e. beneficiaries of the LA process, but also data subjects, i.e. the suppliers of data (Greller & Drachsler, 2012). Data subjects are usually learners, but also teachers, and in many use cases of LA the two types of stakeholder groups will actually be the same.

## 2.4 Objectives & Key Applications

The objectives for using LA are in line with the different views of its stakeholder groups. In general terms, three types of overarching LA goals can be distinguished: goals related to behavioural reactions of learners, behavioural reactions of teachers, and goals to inform the design of learning analytics tools and methods (Dyckhoff et al., 2013). Chatti et al. (2012) identified a set of seven main objectives. These objectives certainly have overlaps and usually a specific application of LA will serve several of them.

- **Monitoring and analysis**: Tracking and checking the learning process, which is then used by teachers or educational institutions as a basis for taking decisions, e.g. on future steps, the design of learning activities, improving the learning environment.
- **Prediction and intervention**: Estimating learners' future knowledge or performance in terms of finding early indicators for learning success, failure, and potential dropouts, to be able to offer proactive interventions and support for learners in need of assistance.
- **Tutoring and mentoring**: Helping learners with and in their whole learning process or in the context of specific learning tasks or a course, providing guidance and advice.
- **Assessment and feedback**: Supporting formative and summative (self) assessment of the learning process, examining efficiency and effectiveness of learning, and providing meaningful feedback on the results to teachers and learners.
- **Adaptation**: Finding out what a learner should do or learn next and tailoring learning content, activities, or sequences to the individual. This idea of carefully calculated adjustments corresponds to the central aim and component of adaptive learning environments and intelligent tutoring systems.
- **Personalization and recommendation**: Helping learners in deciding over their own learning and learning environment, and what to do next by providing recommendations, while leaving the control to the learner.

- **Reflection**: Prompting and increasing reflection or self-reflection on the teaching and learning process, learning progress and achievements made; providing comparison with past experiences or achievements, between learners, across classes etc.

Applications of LA refer to the actual usage of LA in concrete application scenarios. A specific application will certainly pursue one or several of the above indicated objectives and will try to achieve this goal in practical implementation of an LA process. In terms of the key applications of LA, Baker and Yacef (2009) took a highly research- and development-oriented view in their review and highlighted the following application areas: improvement of learner models, improving or uncovering models of a knowledge domain's structure, investigation of pedagogical support to find out which types of support are (most) effective, and finding empirical evidence for refining or elaborating educational theories and phenomena for a better understanding of learning and its influencing factors and as an information source for learning system design. These applications focus first and foremost on questions relevant for educational and LA researchers.

In contrast, Romero and Ventura (2010) took a more education- and practice-oriented position and identified a set of educational tasks that LA may be applied for. Although they partly overlap with Baker and Yacef (2009, see above), these are applications that are particularly relevant for learners, teachers, and educational institutions:

- Providing meaningful information, feedback, and visualisations to support instructors and educational administrators in decision making on instruction and proactive or remedial action
- Providing recommendations to students
- Adapting learning contents, sequences, and interfaces
- Predicting learner performance
- Learner modelling, including e.g. skills, motivation, learning styles
- Detecting undesirable or erroneous learner behaviour
- Supporting the creation of student groups
- Studying the relationships between learners
- Supporting teachers in concept map creation
- Assisting construction and reuse of learning content
- Enhancing educational planning and scheduling

The majority of current LA applications are carried out towards intelligent tutoring system design and research (Chatti et al., 2012). The main objectives followed are thereby monitoring, analysis, and adaptation. The educational tasks that may be supported by LA and are listed above show that beside using LA for understanding what learners do, predicting what they will do or how successful they'll be, and personalising and improving learning experiences, LA can and should have also an important role in terms of supporting higher level educational decisions and transforming the educational landscape. LA may serve a valuable information source for educational administrators and decision makers to shape education and allocate resources to optimise learning and educational results at national and international levels (Ferguson, 2012; Long & Siemens, 2011).

## 2.5 Data in and for Learning Analytics

The data used as a basis for analytics in a LA project first and foremost needs to be educationally relevant; thus, the significance of LA naturally always depends on the educational data available and used for the analytics process. For LA to be successful, an effort needs to be put on identifying which data is available or can be made available that can provide meaningful insights and is in line with the intended objectives (Pistilli, Willis, & Campbell, 2014).

There is a wide range of learning systems and tools available and used in an educational context or for learning purposes (ITS, LMS, concept mapping, social networks), and all of them provide different kinds of data. Data used for learning analytics may come from a single or a variety of sources; it may be structured (like server logs) or unstructured (like forum postings) (Van Harmelen & Workman, 2012). In general, two big categories of data sources can be distinguished (Chatti et al., 2012):

- **Centralized** educational systems, like LMS, provide extensive log data of learner interaction and activities (accessing learning resources, reading, writing, taking tests).
- **Distributed** learning environments provide multiple logs from a range of different sources from formal and informal channels and distributed across space, time, and media. The data thereby my come from formal or informal learning channels.

A challenge when dealing with educational data from distributed learning environments for LA is the issue of data integration from those different sources and potentially represented in different formats. Another issue is the storage of data, since analytics processes by nature use ‘big data’, i.e. large data sets that would not be practicable to deal with for manual analysis (Ferguson, 2012). When considering large data sets, a distinction between extensive and intensive data can be made (Homer, 2013):

- **Extensive data** refers to data that is collected from a large number of participants on a limited number of variables (e.g. data from a widely used massive open online course), resulting in a wide but shallow set of data, which is typically used for traditional data mining techniques.
- **Intensive data,** on the contrary, is data from a relatively small number of participants, but with detailed observations on a large number of variables, thus resulting in a deep but narrow data set. Intensive data usually consists in several traces or logs of data; analysis is done across these different traces.

Extensive and intensive data can meaningfully complement each other, for example for triangulation and validation of results or by using intensive data for deciding on the type of extensive data to be collected (Homer, 2013).

In any case, there is a large variety of learning-related data and traces that result from the use of different pieces of software and media and that potentially may be used for analytics. This is not only confined to traditional learning management systems, but includes a whole range of additional tools. In particular, also massive open online courses, which are emerging rapidly, can be seen as an important field where large-scale educational data can be gathered from and learning analytics can be effectively applied. Beside that, social media, geo-location services, and wearable sensors collecting data of an individuals personal life, activities, physical state etc. are presumed to offer additional data sources that may be exploited for LA in the future (MacNeill et al., 2014). In particular data collected from multimodal interaction and sensing technologies (e.g. heartbeat, speech, gesture) is considered as having potential to provide additional insights into students' learning (Blikstein, 2013). In general, the integration of data from multiple sources can improve the accuracy of the models established for learners and thus, helps improving action following (e.g. adaptation, intervention by teacher) from that (Papamitsiou & Economides, 2014).

Deciding on the kind of data to be actually captured and the information to be extracted is key in the LA process. The choice of data used as predictors and indicators immediately influences the type, quality and accuracy of the analysis. The selection of data also involves some sensitivity, in terms of the decision whether information – usually held by an educational institution – about social background etc. should be used for analytics. This very much depends on the aims and questions addressed by analytics.(Van Harmelen & Workman, 2012). In general, three broad types of indicators can be distinguished (Brown, 2013):

- **Dispositional indicators** are factors that the learner brings to the learning context and are available before the learning episode begins. Examples are age, gender, previous learning experiences etc. Many of them are factual, readily quantifiable and available in the student model of learning systems or for an educational institution.
- **Activity and performance indicators** refer to data that learners produce as they are engaging in learning activities and making their way through a course. Examples are the number of logins, time spent, number of discussion posts etc.
- **Student artefacts** consist in data resulting from learners' actual work in terms of the products of the learning process, like essays, blog posts, discussion forum contributions. Analysing such artefacts can provide information on the mastery or competence of learners.

The data and indicators that can be selected for data collection necessarily are based on the data that is available from the learning environments and applications. Data tracking occurs without any extra manual effort by the learner. Thereby it is of course crucial that learners are aware that their data and activities are logged (Duval, 2011).

Dyckhoff et al. (2013) conducted a comprehensive review on state of the art of LA and collected about 200 indicators currently used LA (e.g. number of threads per student, number of participants per group, clusters of student who made a specific mistake etc.). They categorised the indicators according to the different perspective one may have on the data (individual learner, group, course, content, and teacher). Additionally, the origin of data (data sources) was mapped to six categories: student generated data, context/local data, academic profile, evaluation, performance, and course meta-data. This review showed that a large part of the data used in current LA tools is basic usage data (i.e. activity and performance indicators) of learners engaging with a single learning environment. The authors conclude that in order to be able to answer more complex, highly relevant research questions that educational practitioners have in mind, a greater emphasis needs to be put on high-level indicators and teachers should optimally be actively involved in the definition and design of relevant indicators.

Data collection in LA is not confined to the pure capturing of learners' traces via different indicators, but may also consist in the combination of data from different sources, as indicated earlier. This is done in the data pre-processing and aggregation phase of LA, i.e. the datasets are merged and transformed into a suitable format for further analytics processing. It has to be taken in mind that in many cases the data is collected by systems for a different use and without necessarily intending analytics purposes. One of the most difficult challenges with respect to data for learning analytics consists in establishing a common meaning to the data and to create interoperability in analytics, to make information flows and analytics procedures more efficient, timely, and effective (Cooper, 2014). Establishing a generic data model and standards for collecting data would considerably increase interoperability in LA and reduce development costs of tailoring analysis tools to learning systems, a generic (del Blanco, Serrano, Freire, Marítinez-Ortíz, Fernández-Manjón, 2013; Fortenbacher, Klüsener, & Schwarzrock, 2014).

A relevant issue when considering data for LA is that of privacy and ownership. LA strongly relies on data about learners, and to be able to test and evaluate LA methods, access to appropriately large available datasets are needed. Thus, there are recent attempts and initiatives on making educational data more open. PSLC DataShop[3] (Koedinger et al., 2010), for example, is a repository of data from a variety of coursed. Key aspects of such initiatives on opening access to learning data are anonymisation of data according to legal requirements, a unified documentation and data format, as well as data policies regulating use and sharing of such data sets (Greller & Drachsler, 2012).

At the moment, centralized web-based learning systems are still the most common data source for LA, and the kind of data used are largely activity and performance indicators (Chatti et al., 2012; Dyckhoff et al., 2013). With analytics approaches becoming more mature there will be increasingly a convergence of different data sets, i.e. institutional data sources, data that learners share voluntarily, and a whole variety of other relevant signals and data available (e.g. from social media etc.) (Buckingham Shum, 2011).

## 2.6 Visualizations

LA is not only about collecting and analysing educational data, but also feeding back and making use of the results is essential. The results and inferences of LA are usually used in the following ways: either the information is fed into adaptation and recommendation mechanisms, or it is reported back to

[3] https://pslcdatashop.web.cmu.edu/

the learner, teacher, or other stakeholders to empower and support the teaching and learning process. In the latter case, visualisations play an essential role. Learning analytics, aside from helping teachers and administrators to study learners, are considered to have a great potential of providing mirrors for learners and helping them to become more reflective and less dependent, and to foster 21st century skills (Shum, 2011). So, visualisations are used reflectively, to promote reflection and metacognitive action, and they may serve supporting on-the-spot decision making by learners or teachers (Bull et al., 2013).

Different disciplines contribute to visualisation: graphics design and information visualisation, cartography, computer science, psychology, statistics. The fine-grained statistics available from LA are oftentimes too voluminous and cumbersome to inspect or too complicated and time consuming to interpret. Visualizations can help people to understand and analyse the data (Romero & Ventura, 2010). In fact, visualisations provide the opportunity of giving an overview on large amounts of data, which would otherwise be hard to take in (Bull et al., 2013). In this way, visualisation of analytics and analytics results can significantly support the adoption of analytics approaches (by non-experts) (Van Harmelen & Workman, 2012).

Suitable visualisations play an essential role in making big sets of learning-related data and LA results better understandable, to gain an insight in the learning and teaching process and the interrelation between teaching and learning. This is, in fact, a prerequisite for achieving the overarching goal of LA in terms of gradually improving teaching and learning processes (Chatti, Dyckhoff, & Schroeder, 2012).

Visualizations make LA results actionable, i.e. they enable teachers, mentors, learners to take appropriate decisions and action (Brown, 2013). Thereby, visualizations will differ in the way results are displayed (chart or diagram type) and the way results are presented for different stakeholders.

Visualisations of learning traces are called learning dashboards and are commonly applied in LA (Verbert, Duval, Klerkx, Gofaerts, & Santos, 2013). They enable teachers and learners to get an overview of their activities and how they compare to others. Different approaches of dashboards exist. “All-at-one-time” dashboards represent different visualisations with different aspects of information side-by-side. Other dashboard approaches start with one visualization and enable the user to access further information and detail from there (Brown, 2013). A variety of dashboard applications have been developed recently; the learning dashboard approach is considered to have very good potential to foster awareness, reflection, sense making and, in the end, improve learning. However, evaluation of the actual impact of using them is difficult and rarely documented in current state of the art (Verbert et al., 2013).

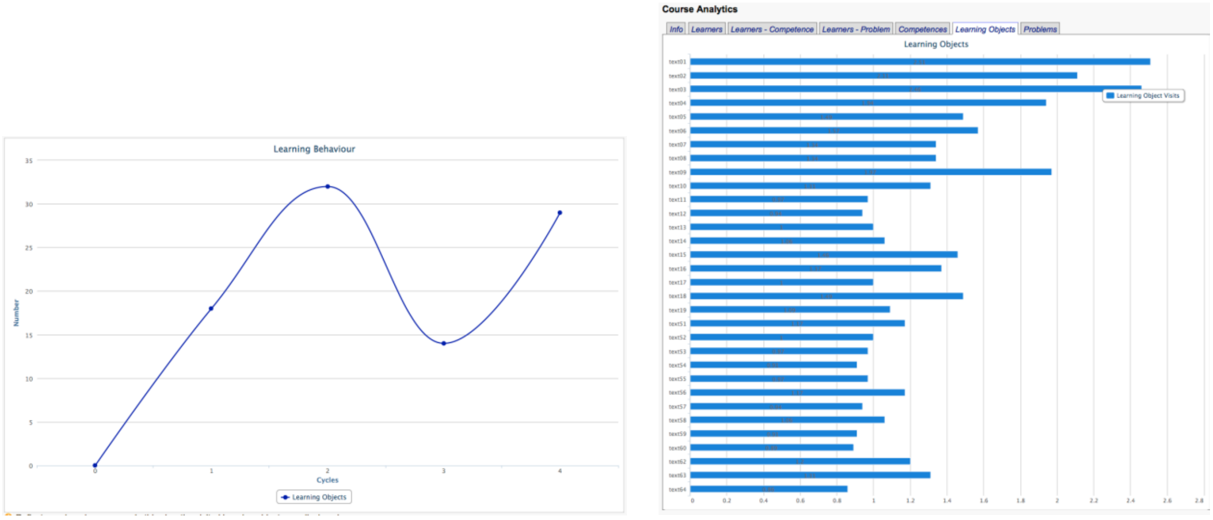


**Figure 6:** Example LA visualizations on interaction data (learning object visits).

Visualisations may just show activity and interaction data (e.g. time spent on pages, learning objects visited; see Figure 6 for examples), but may also depict the inferences dawn from that data through analytics to learners' understanding, their competencies etc. (see Figure 7 for examples). LA visualizations for learners (or other stakeholders) are therefore related to the topic of Open Learner Models (OLM), i.e. the idea of opening up the learner model to the user to support reflection on and awareness of learning, and dynamically update that information for a deeper understanding of the learning and teaching process (Bull & Kay, 2010). This is similar to the goals of LA, but while LA visualizations today oftentimes tend to be confined to the illustration of activity and interaction data, OLM focuses on the representation of inferences drawn from that data in terms of learners' skills, knowledge, affective states (Bull, Kickmeier-Rust, Vatrapu, Johnson, Hammermueller, Byrne et al., 2013).


**Figure 7:** Example LA visualizations on inferences drawn on learners' competences (skill meter and radar plot).

With the evolution of LA, the need of responding to more complex educational questions and making inferences to competence and mastery, visualisations in LA are also developing more and more towards including or representing this kind of information (e.g. Grann & Bushway, 2014). More sophisticated visualisations may facilitate finding positive evidence of their impact on learning, e.g. on competence scores and persistence rates, as well.

A more in-depth review about educational data visualisation approaches and open learner modelling and their relevance and implications for LEA's BOX will be provided in a separate deliverable (D4.1 in M18).

# 3. LEARNING ANALYTICS METHODOLOGIES

LA provides results that my inform teachers, students, or administrators and provide a basis for appropriate action and intervention on course-level or institutional level (Van Harmelen & Workman, 2012). The kind of analytics methodologies employed in a concrete LA project depends on the addressed stakeholders and their objectives (see sections 2.3 and 2.4) and on the kind of data collected (see section 2.5). This section gives an overview of common methodologies, i.e. methods for analyzing educational data (section 3.1) and tools implementing those methods (section 3.2).

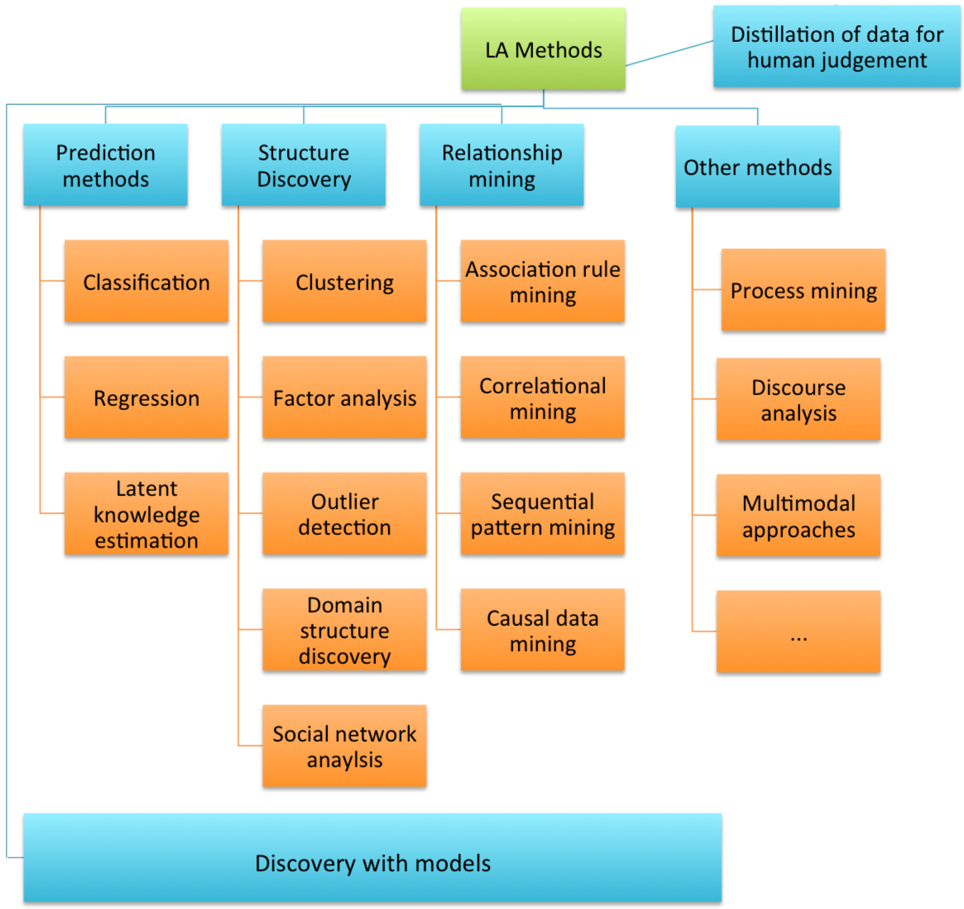


**Figure 8:** Overview of LA methods.

# 3.1 Analytics Methods

In LA different methods are used to extract meaningful patterns from educational data. The techniques actually used in a concrete application scenario will depend on the objectives of the analytics tasks, but also on the kind of data collected. Baker and Siemens (in press) consider methods from data mining and analytics, in general, as well as psychometrics and educational measurement as the main sources of inspiration for LA methods and tools, and they provide a systematic overview of the key methods currently applied in LA, which fall in five main classes: prediction methods, structure discovery, relationship mining, discovery with models, and distillation of data for human judgment (see also (Baker & Inventado, 2014). Below a summary on the most important methods applied for LA is given (based on Baker & Inventado, 2014; Baker & Siemens, in press; Romero & Ventura, 2013). An overview is presented in Figure 8.

## 3.1.1 Prediction Methods

These are the most popular methods in LA and essentially aim at developing a model to predict or infer a certain variable (predicted or dependent variable, e.g. mark, performance score) from a combination of other indicators (predictor variables or independent variables) of the educational data set (Baker & Siemens, in press). According to a recent review of LA approaches (Papamitsiou & Economides, 2014), among these, classification is used most often. For establishing this kind of model, knowledge of the predicted variable for a restricted data set is necessary to be used as so-called ground-truth for the prediction model. Prediction models are created for variables that are not feasible or possible to collect in real-time, or to forecast future events or performances. Prediction models are also relevant when analysing which indicators are relevant in predicting another variable.

Common prediction methods are:

- Classification: for prediction of binary or categorical variables, e.g. decision trees, step regression
- Regression: for prediction of continuous variables, e.g. linear regression
- Latent knowledge estimation: for assessing learner knowledge or skills; i.e. students' knowledge or skills based on their performance, their patterns of correctness on that knowledge/skills, e.g. Bayesian Knowledge Tracing

## 3.1.2 Structure Discovery

Algorithms of structure discovery aim at detecting structure in educational data without an a priori assumption of what should be found (Baker & Siemens, in press). Examples are grouping of students, student actions, and system features. This is in contrast to prediction methods, where the predicted/dependent variable is known.

Methods of this type are:

- Clustering: splitting data set into clusters grouping data points together, i.e. identifying groups of instances that are similar
- Factor analysis: finding dimensions of variables grouping together
- Outlier detection: discovering data points that are significantly different from the rest of data
- Domain structure discovery: deriving the structure of knowledge in an educational domain

Another method from this class, which is quite popular in LA, is social network analysis (SNA) (Bakharia & Dawson, 2011). It allows analysing relationships and interactions between learners in terms of collaboration and communication activities, information exchange etc. SNA uncovers the patterns and structure of interaction and connectivity, which can then be visually illustrated. SNA draws on concepts from graph and structural theory. It allows assessing network properties, like density, centrality, connectivity, betweenness, and degrees. This means, SNA provides the possibility of quantifying the social interactions and network between learners, to identify learners that are very important, represent 'hubs', or are in isolation (Chatti et al., 2012).

### 3.1.3 Relationship Mining

The aim of this group of methods is to find out relationships between variables in a large set of variables and how strong those relationships are (Baker & Siemens, in press). These methods provide instruments to uncover meaningful, but potentially unexpected relationships between variables. An example is to use relationship mining techniques to find patterns of learning behaviour that are associated with successful student performance.

There are four methods of relationship mining commonly applied in LA:

- Association rule mining: finding if-then rules
- Correlation mining: finding positive or negative linear correlations
- Sequential pattern mining: finding temporal associations between events
- Causal data mining: finding out whether one observation is the cause of another

### 3.1.4 Discovery with Models

This class does not denote a specific group of techniques but refers to the general approach of using the results of one analytics method within another analysis (Baker & Siemens, in press). Popular ways of doing this is for instance the use of a prediction model within another prediction model or using a prediction model within a relationship mining analysis, but there are a variety of other ways for conducting discovery with models.

### 3.1.5 Distillation of data for human judgement

This is an approach that is common in LA, in a narrower sense, but not considered as a traditional method of EDM, since it consists in providing teachers immediate access to reports and visualisations of the learner data, for their interpretation and judgement, and to support decision making and pedagogical action (Baker & Siemens, in press). Examples are learning curves or heat maps (Serrrano, Marchiori, del Blanco, Torrente, & Fernandez-Manjon, 2012). Visualisations play a critical role in this context, since the LA data needs to be represented in an appropriate way to support educators in understanding and judgement.

### 3.1.6 Other methods

**Process Mining**

Process mining methods extract process-related information from event logs to derive a visual presentation of the whole process. Three subfields can be distinguished: conformance checking, model discovery, and model extension. Educational process mining (EPM) aims at (i) constructing complete and compact educational process models that are able to reproduce all observed behaviour (process model discovery), (ii) checking whether the modelled behaviour (either pre-authored or discovered from data) matches the observed behaviour (conformance checking), and (iii) projecting information extracted from the logs onto the model, to make the tacit knowledge explicit and facilitate better understanding of the process (process model extension (Trcka & Pechenizkiy, 2009, p. 1114).

**Discourse Analysis**

Discourse analysis, in general, serves identifying the type and content of conversations, their communicative functions, the quality and purpose of discussions and collaboration (e.g. Arvaja & Pöysä-Tarhonen, 2011; Mercer, 2010). Discourse analysis may be done using different kinds of approaches, like human-led qualitative methods or quantitative (nature, patterns, and quality of interactions), usually computer-supported methods (counts of specified features of discourse), sophisticated natural language processing and text mining, like text categorisation, concept extraction, semantic sensing, document summarisation (e.g. Mercer, 2010; Despotakis, Dimitrova, Lau, Thakker, Ascolese & Pannese, 2013).

Discourse analysis originally focused on spoken discourse, like classroom talk (Mercer, 2010), but subsequently was also applied to textual discourse in online learning, and – with the formalisation of coding and interpretation schemes for computational modelling and NLP – opened up the application for LA purposes (Buckkingham Shum, 2011; De Liddo, Buckingham Shum, Quinto, Bachler, & Cannavacciuolo, 2011). Discourse analysis thereby basically consists in analysing written evidence of learning activities and online communication, like discussion forum posts etc..

More interest has been dedicated to discourse as an important indicator for learning and thus, on focusing on analysing learners' discourse for LA purposes, to identify where and how learning happens and better understand the learning process (De Liddo et al., 2011). This is inline with the aim of more sophisticated analytics to establish a deeper understanding of learning, in addition to more generic analytics harvesting quantitative data (Ferguson & Buckingham Shum, 2011). Social network analysis, for example, which provides information on the structure of interaction between learners, can be meaningfully complemented by discourse analysis, which may give additional insight on the kind and quality of interaction. While SNA reveals detailed insight on the structure of interactions between learners, it does not provide an understanding of the content or nature of communication. Discourse analysis may help to better understand the interaction patterns observed. One example method that is used in discourse-centric LA is Epistemic Network Analysis (ENA), which actually adapts tools of social network analysis and originally was devised and applied in the context of virtual games (Shaffer et al., 2009; Shaffer & Arastoopour, 2013). ENA can be used to analyse the connections between discourse elements, their pattern, and their development.

**Multimodal approaches**

Multimodal technologies are starting to be recognised as a way of collecting data for the purpose of learning analytics that may meaningfully complement data from educational tools and learning systems, to provide additional insight on learning. Natural rich modalities of interaction and communication, like speech, writing, gestures, face expressions, gaze, are collected during computer-supported or interpersonal learning activities. Novel data collection and sensing technologies (e.g. wearable sensors, cameras, digital pens) enable collection of comprehensive data on human activity is possible. This multimodal interaction data is analysed via different methods, like gesture sensing, face recognition, eye tracking, speech recognition etc. Multimodal learning analytics is denoted as "a set of techniques that can be used to collect multiple sources of data in high-frequency (video, logs, audio, gestures, biosensors), synchronize and code the data, and examine learning in realistic, ecologically valid, social, mixed-media learning environments" (Blikstein, 2013, p. 105). The idea behind a multimodal approach is to combine different logging and sensor information for a naturalistic assessment of learning experiences and analysis of learning settings (Shoukry, Göbel, & Steinmetz, 2014). With multimodal approaches evolving from discontinuous and obtrusive detection and analysis to the possibility of continuously and unobtrusively collecting data (e.g. remote eye trackers instead of head-mounted eye trackers) and directly drawing inferences to the learning experience; this, in turn, paves the way for the use of multimodal techniques for LA projects focusing on the present and aiming at a continuous feedback loop of LA results to learning and teaching (Bahreini, Nadolski, & Westera, 2014).

## 3.2 Analytics Tools

A variety of tools exist for applying LA methods and for visualising data. The tools come from the academic and the commercial sector (with tools from the latter usually originating from the academic sector). Increasingly companies offer 'analytics as a service' products for a fee (MacNeill et al., 2014); these are usually not confined to educational applications, but offer broader functionality, oftentimes with a special focus on business analytics.

LA tools implement different LA methods (as outlined in section 3.1 above), and provide support in data pre-processing, validation of models, and visualisation of data. While proprietory and marketed LA software tools are usually black boxes that do not provide detailed insight or changes to the implemented methods and algorithms, open platforms allow scrutinizing the methods used and developing add-ons to extend analytics (Baker & Siemens, 2014).

The LA tools available have different target audiences and applications; they are geared to application for research purposes, i.e. by researchers and experts, or to application in educational practice, i.e. by instructors, learners and educational institutions. In line with these, LA tools apply and provide different approaches. Although some tools find already application in educational practice, many of the existing tools are very complex and are solely used for the purpose of LA research and development, but do not appropriately fit the needs of educators (Romero & Ventura, 2010; Siemens, 2012). In general, LA tools providing visualisations with advanced and more powerful functionality are commonly more difficult to learn, but allow exploring more complex questions like investigating the relationships between different variables (Buckingham Shum, 2012). Often only technically skilled or expert users can utilise these LA tools (Dyckhoff, 2011).

Overall, there is a wide array of tools that may be used to support the LA process in some way and an exhaustive review is hardly possible. Collections of currently available tools have for example been presented by Kraan and Sherlock (2013), Dornan (2012), Romero and Ventura (2013), and Schneider (2014). In the following subsections, different categories of LA tools will be described and for each type a (non-exhaustive) list of tools is provided that are available now. (Note: The different categories are not mutually exclusive and some tools provide functionality actually related to several of these categories.) For illustration selected example screenshots are presented, as well. In general, not all tools are relevant to all application areas of LA, and different tools are designed to deal with a different range of data sources (Kraan & Sherlock, 2013) and implement different methods.

### 3.2.1 Tools for Extraction, Transformation, Loading

Before conducting data analysis by applying an LA method and subsequently visualising its results, the data must be acquired, integrated, cleansed, and stored in an appropriate way (Dornan, 2012). While some LA tools include functionality for pre-processing data, there are also tools available specifically for that purpose – so called ETL (Extraction, Transformation, Loading) tools. An ETL tool extracts data from different sources, it transforms the data for analysis and reporting, synchronises data from different databases, cleanses data to remove errors/noise, and loads data into a data warehouse (Rayón Jerez, 2014). See Table 2 for a list of ETL tools and Figure 9 for an example screenshot.

**Table 2:** ETL tools.

| Tool name | Reference | Description |
|---|---|---|
| Pentaho Data Integration | http://www.pentaho.com/product/data-integration | - visual tool for preparing and blending big data, to make it 'analytics ready'<br>- originally an open source tool, but meanwhile part of a commercial business analytics suite |
| Talend Data Integration | https://www.talend.com/products/data-integration | - extensible set of tools to access, transform and integrate data for operational and analytical data integration needs<br>- open source tool; additional functionality (e.g. teamwork functions, clustering) available as subscription |
| Yahoo! Pipes | https://pipes.yahoo.com/ | - web application providing a graphical user interface for aggregating and mashing up we-based data |

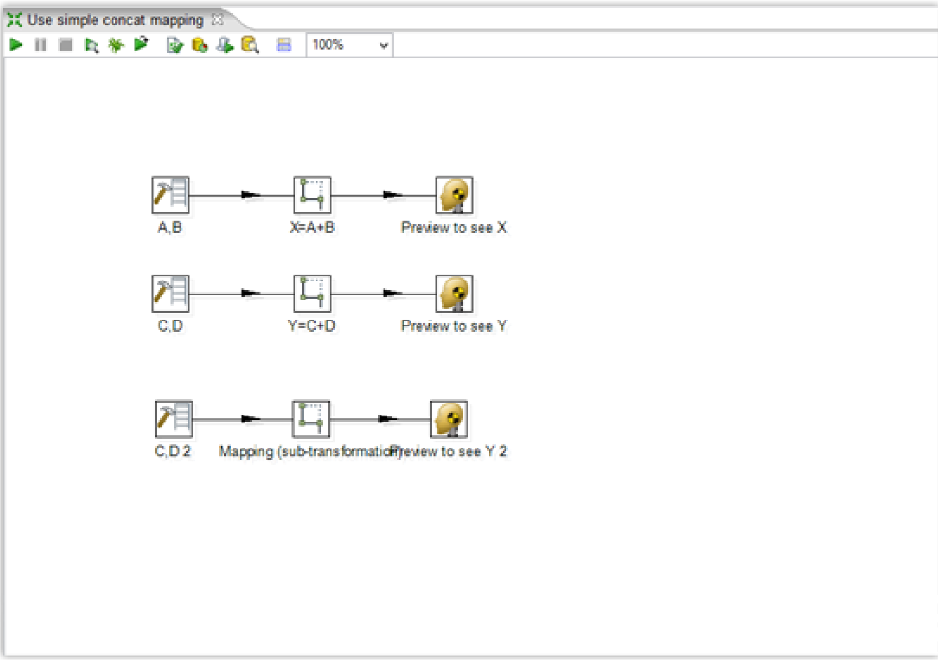


**Figure 9:** Screenshot of Pentaho Data Integration (source: http://community.pentaho.com/projects/data-integration/).

### 3.2.2 Web Analytics Tools

Web analytics usually means fact based reporting on an organization's web presence in terms of reporting page visits, visit length, search terms, visitors' location etc., or may mean analyzing data related to an organization, like the public sentiment on the organization or its product (Kraan & Sherlock, 2013). Web analytics tools can also be used to set up basic analytics for open online courses. "The simplest learning analytics are those that count the hits and time spent on certain content pages" (de Waard, 2012). This is, of course, simple reporting of activity data, but it may nevertheless provide some first insight on the learning process in terms of the visits, paths, time spent, platform used, location etc. Most web analytics tools require inserting a JavaScript snippet into the webpage to be analyzed, to enable user tracking (Kraan & Sherlock, 2013). There are numerous web analytics tools available, many of them are commercial, but minimal services are oftentimes free. See Table 3 for a list of web analytics tools and Figure 10 for an example screenshot.

**Table 3:** Web analytics tools.

| **Tool name** | **Reference** | **Description** |
|---|---|---|
| Google Analytics | http://www.google.com/analytics/ | - commercial web analytics service offered by Google<br>- most widely used statistics service<br>- stats can be viewed through a web browser |
| Open Web Analytics (OWA) | http://www.openwebanalytics.com/ | - open source web analytics software to track and analyse use of websites and web application |
| Piwik | http://piwik.org/ | - open source web analytics platform<br>- additional features supporting analysis of data collected (e.g. annotation, setting goals) |

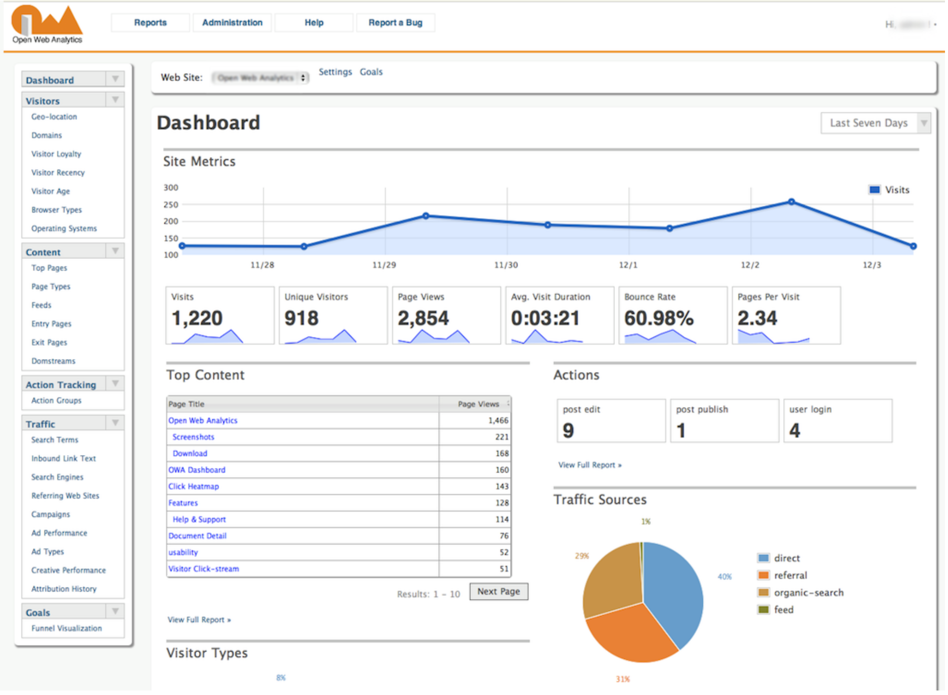


**Figure 10:** Screenshot of main analytics dashboard from OWA (source: http://www.openwebanalytics.com/)

### 3.2.3 Business Intelligence Tools

LA is usually considered as having its roots in the field of business intelligence. Business intelligence (BI) or business analytics provides computational tools that focus on improving organizational decision making (Buckingham Shum & Ferguson, 2012). BI tools are designed to be general purpose and the data sources that can be included with these tools is very broad. This provides the opportunity of using BI tools also for LA purposes. Sometimes even some education specific features are built in, but due to the generic orientation of BI tools, some customization will most likely be necessary (Kraan & Sherlock, 2013). There is a wide range of business analytics tools available, a selection is listed in Table 4; an example screenshot is shown in Figure 11.

**Table 4:** BI tools.

| Tool name | Reference | Description |
|---|---|---|
| Metrics That Matter | http://www.executiveboard.com/exbd/human-resources/metrics-that-matter/index.page | - human capital analytics and business intelligence system<br>- cloud-based talent analytics as a complement to an LMS for capturing and reporting the effectiveness of learning programs |
| Microsoft BI suite | http://www.microsoft.com/powerbi/ | - cloud-based BI environment<br>- Excel can be used to analyse and visualise data, thus supporting nearly all steps of the analytics process |

| | | |
|---|---|---|
| Pentaho | http://www.pentaho.com/product/business-visualization-analytics | - business analytics suite providing analytics from basic reports to predictive modelling |


**Figure 11:** Screenshot of dashboard on business results from Metrics That Matter (adapted from KnowledgeAdvisors, 2009).

### 3.2.4 Information Visualization Tools

The discipline of information visualization is concerned with creating "visual artifacts aimed at amplifying cognition" (Mazza, 2009, p. 8), i.e. information visualization takes data and represents it in a visual format to be easily process- and understandable by humans. Kraan and Sherlock (2013) highlight that here is a relation to other types of tools, since the data needs to be collected and prepared first. Exemplary tools are listed in Table 5; an example screenshots is given in Figure 12.

**Table 5:** Information visualization tools.

| Tool name | Reference | Description |
|---|---|---|
| Cytoscape | http://www.cytoscape.org/ | - open source software platform for visualising complex networks<br>- originally designed for biological research, but now a general platform for network analysis and visualisation<br>- basic feature set for data integration, analysis, and visualisations; additional features via Apps |

| | | |
|---|---|---|
| Gephi | https://gephi.github.io/<br><br>Bastian, Heymann, & Jacomy (2009) | - open source software for exploring and manipulating networks<br>- visualisation and exploration of all kinds of networks and complex systems, dynamic and hierarchical graphs |
| Tableau | http://www.tableausoftware.com/ | - free analytics visualisation software to create interactive visualisations on existing data sets and to publish them to the web<br>- uses VizQL, a visual query language |

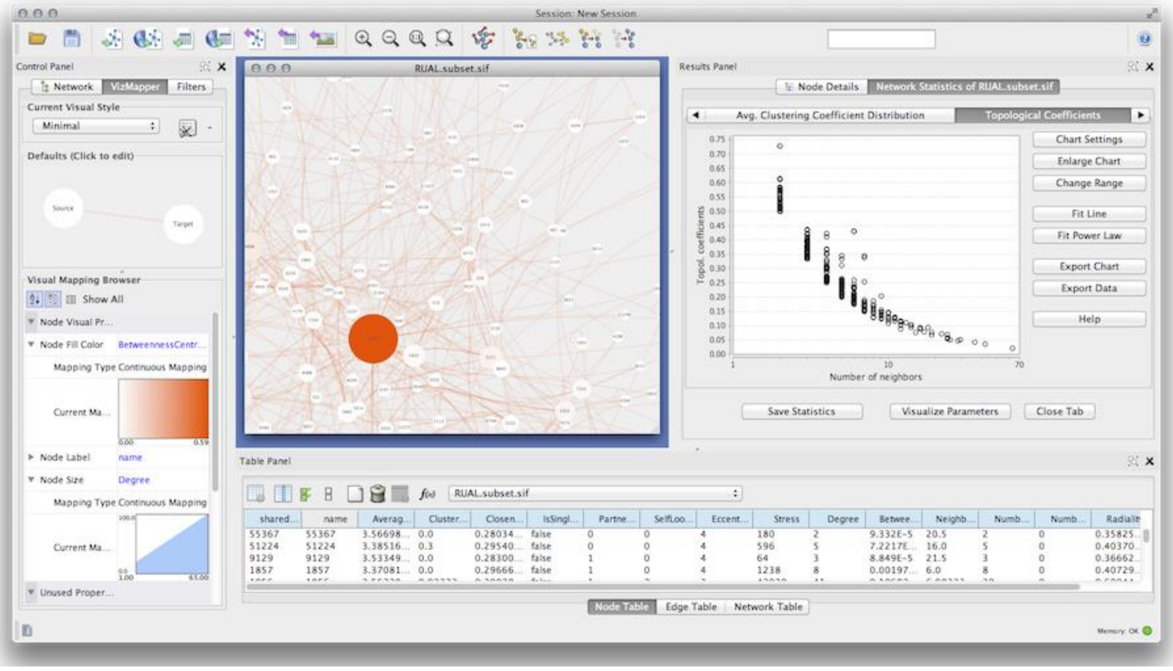


**Figure 12:** Screenshot of Cytospace with sample visualization (source: http://www.cytoscape.org/).

## 3.2.5 Social Network Analysis Tools

Social network analysis (SNA) tools allow analyzing the relationships/interactions between individuals and to visualize the resulting social networks. This category of tools can be considered a special type of information visualization tools (cf. section 3.2.4). Numerous tools are available for interactive visualisation and analytics of social networks (Buckingham Shum, 2012); a prominent example in the LA field is SNAPP (Social Networks Adapting Pedagogical Practice). It especially aims at providing teachers the possibility to evaluate learner behaviour against learning objectives as a basis for interventions. The visualisation of the social network of a class or group of learners allows identifying students that are disconnected from the network. Based on the SNA results, isolated students can be supported and encouraged, groups reorganized, and educator effort refocused (Bakharia & Dawson, 2011). SNAPP and other tools relevant for LA purposes are listed in Table 6, a screenshot is presented in Figure 13.

**Table 6:** SNA tools.

| Tool name | Reference | Description |
|---|---|---|
| NodeXL | http://nodexl.codeplex.com/ | - Excel add-on for creating and exploring network graphs<br>- Integrates common network metrics calculations (e.g. centrality, density) |
| NetMiner | http://www.netminer.com/ | - commercial software tool for exploratory analysis and visualization of large network data based on SNA |
| Pajek | http://pajek.imfm.si/doku.php<br><br>de Nooy, Mrvar, & Batagelj (2011) | - free Windows program for analysis and visualisation of large networks |
| SNAPP | http://www.snappvis.org/<br><br>Bakharia & Dawson (2011) | - provides SNA and network diagrams of the activities (forum posts and replies) in a discussion forum hosted by an LMS<br>- especially developed for educational applications |

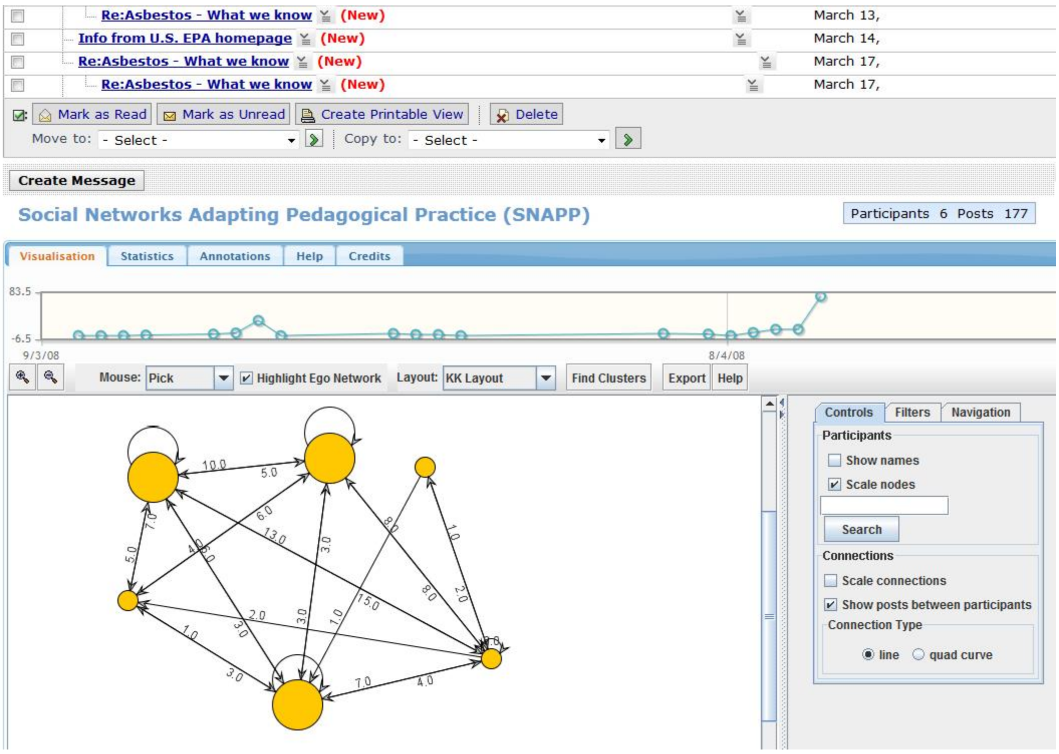


**Figure 13:** Screenshot of SNAPP for visualizing students' discussion forum activity (source: http://www.snappvis.org/).

### 3.2.6 Text analysis tools

There are different tools available to support text or content analysis. Content analysis refers to techniques of analyzing texts and communication data, aimed at reducing large texts and qualitative data into a form condensed to its essential content. This kind of analysis and tools are not limited to learning data, but in principle, to all sorts of texts, like reports, web pages etc. or recorded communication, such as interview transcripts, discourses, observation protocols, social digital traces etc. (e.g. Mayring, 2000; Stemler, 2001). This category of tools include software for supporting (semi-) manual analysis, as well as application for automatic analysis (e.g. Alexa & Zuell, 2000). A selection of text analysis tools is presented in Table 7; a screenshot of one of these tools is depicted in Figure 14.

**Table 7:** Text analysis tools.

| Tool name | Reference | Description |
|---|---|---|
| Leximancer | http://info.leximancer.com/ | - allows to summarize and navigate large text data (e.g. a wiki site)<br>- presentation of concepts found in a text through various visualization tools |
| TagHelper | http://www.cs.cmu.edu/~cprose/TagHelper.html<br><br>Rosé et al. (2008) | - set of tools for automatic text analyses using classification technology and linguistic pattern detection<br>- used for analyses of texts in education; has also been integrated into learning systems and used as instructional tool<br>- built on top of the Weka toolkit (cf. Table 9) |
| Wordle | http://www.wordle.net/ | - web-based tool simple word graphics<br>- generation of word clouds from given texts |

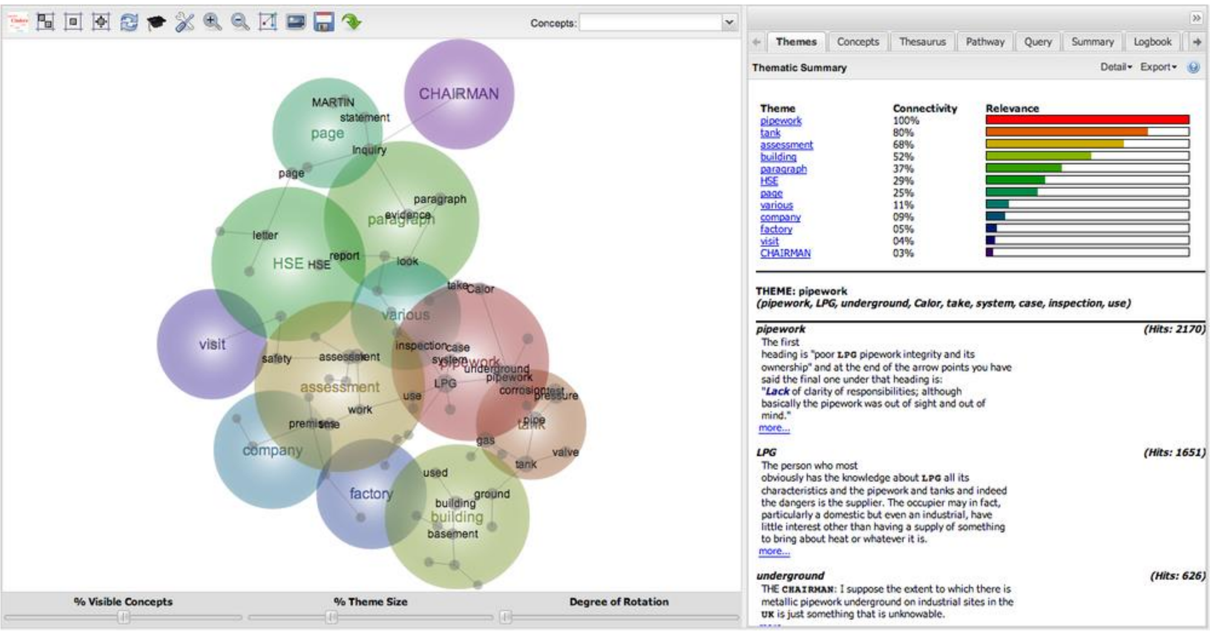


**Figure 14:** Screenshot of Leximancer for visualizing students' discussion forum activity (source: http://info.leximancer.com/).

## 3.2.7 General Purpose Analytics Tools

General statistical software environments have advanced and powerful functionality. They provide comprehensive statistical modeling and analysis methods and usually also visualization and are suitable for building analytic models for the purpose of LA. Open source tools enable the development of add-ons to increase functionality to address specific research needs. Table 8 lists the three most popular statistical computing environments; Figure 15 presents a screenshot of one of these for illustration.

**Table 8:** General purpose analytics tools.

| Tool name | Reference | Description |
|---|---|---|
| R | http://www.r-project.org | - open source statistical computing environment and statistical programming language for statistical and advanced analytics<br>- many data mining algorithms included in standard distribution; numerous additional specialised packages/libraries available for all sorts of functions, to help with data analysis and presentation relevant for LA |
| SAS (Statistical Analysis System) | http://www.sas.com/ | - software suite for advanced analytics, business intelligence, data management, and predictive analytics |

| | | |
|---|---|---|
| | | - provides functionality for ETL, analytics, and visualisation<br>- SAS Enterprise Analytics for Education as contextualized product for the education market targeted at supporting academic analytics for educational institutions |
| SPSS (Statistics Package for the Social Sciences) | http://www-01.ibm.com/software/analytics/spss/ | - general statistics package<br>- extensive functionality for statistical analysis and modelling |

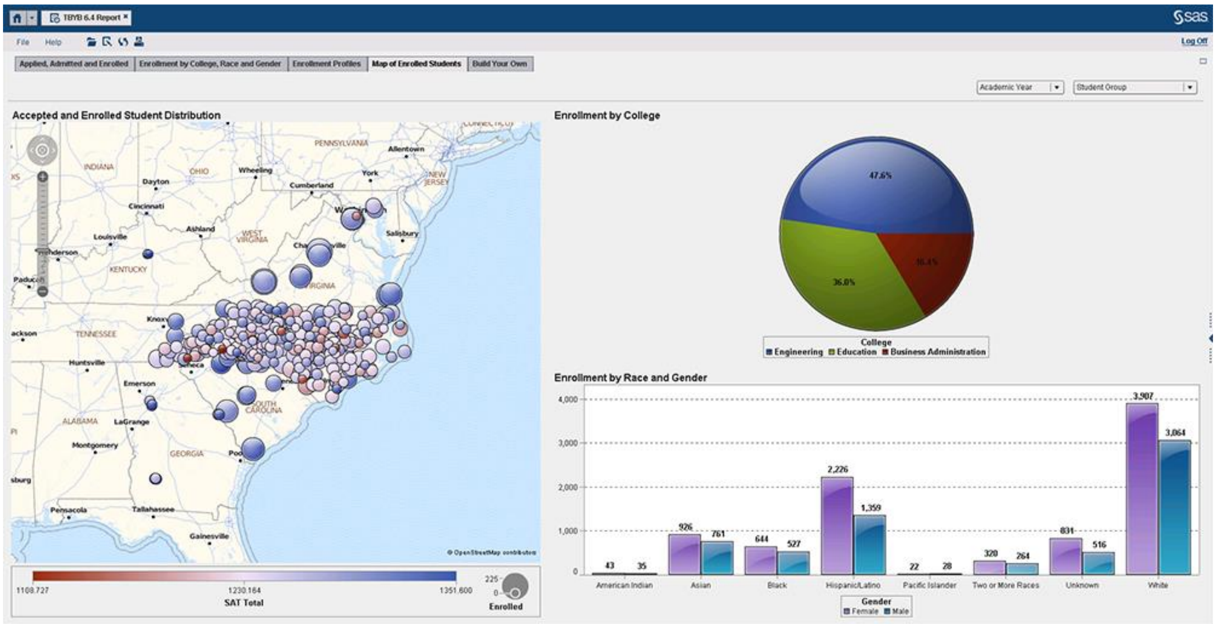


**Figure 15:** Screenshot of enrollment analysis in SAS Enterprise Analytics for Education (source: http://www.sas.com/en_us/industry/higher-education/enterprise-analytics-for-education.html).

### 3.2.8 DATA MINING TOOLS

Data mining refers to the discovery of information from data sets and is also denoted as knowledge discovery in databases or machine learning. There are several software tools that provide extensive data mining applications and allow creating complex analysis models by applying the methods summarized in section 3.1. There are tools that simple implement the algorithms, but usually they provide more comprehensive functionality for data clean-up, visualizations and other steps of the data analysis process (Kraan & Sherlock, 2013). These tools are designed for power and flexibility in carrying out analytics, rather than for simplicity (Romero & Ventura, 2013). Table 9 presents a set of well-known data mining tools; Figure 16 presents an exemplary screenshot.

**Table 9:** Data mining tools.

| Tool name | Reference | Description |
|---|---|---|
| Coron | http://coron.loria.fr/<br><br>Kaytoue et al. (2010) | - free data mining toolkit designed specifically for itemset extraction and association rule generation<br>- originally designed for mining biological cohorts, but generalizable to any kind of database<br>- includes components for pre- and post-processing and visualisation, covering also visualisation of formal concept lattices (see section 8.2) |
| KEEL (Knowledge Extraction based on Evolutionary Learning) | http://keel.es/<br><br>Alcalá-Fdez et al. (2009, 2011) | - open source software tool to assess evolutionary algorithms for data mining problems including regression, classification, clustering, pattern mining etc.<br>- contains a collection of classical knowledge extraction algorithms, preprocessing techniques etc. |
| RapidMiner | http://rapidminer.com/ | - commercial environment for machinge learning, data mining, text mining, predictive analytics, business analytics<br>- methods for data integration, data transformation, data modeling, and data visualization |
| Weka | http://www.cs.waikato.ac.nz/ml/weka/index.html<br><br>Witten, Frank & Hall (2011) | - open source software providing a collection of machine learning algorithms for data mining tasks<br>- contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization<br>- includes advanced text mining features<br>- a number of its data mining tools have been incorporated in other tool, e.g. in RapidMiner |

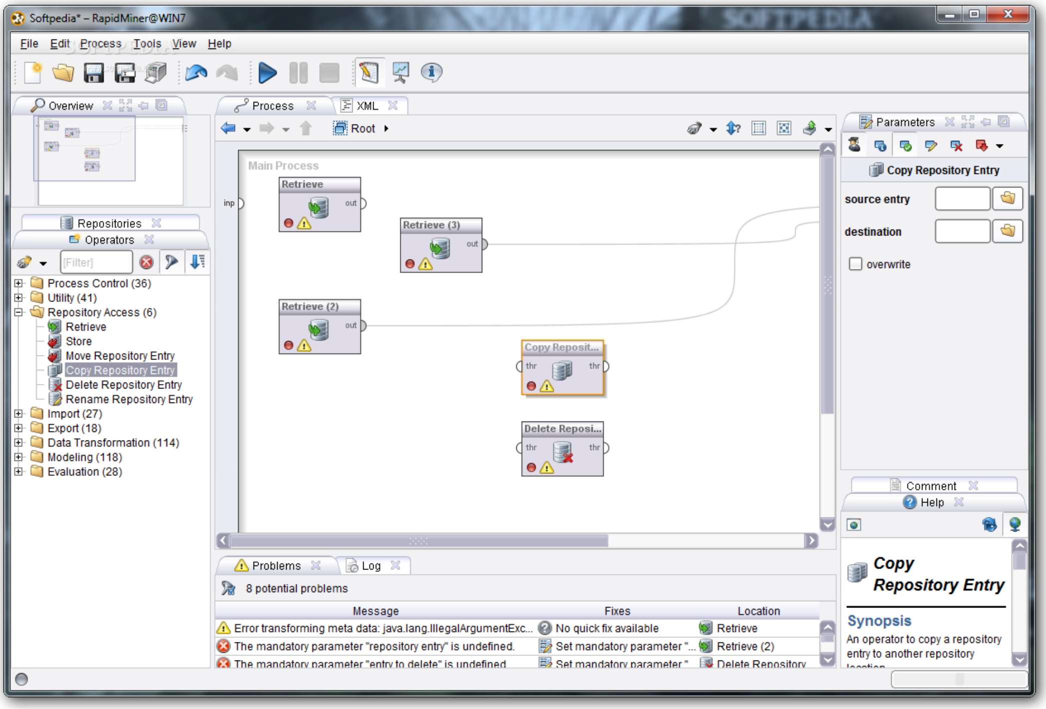

**Figure 16:** Screenshot of RapidMiner
(source: http://www.softpedia.com/get/Others/Miscellaneous/RapidMiner.shtml).

### 3.2.9 Special purpose learning analytics tools

In addition to more general purpose tools for analytics, which usually provide analytics, statistical modeling, and visualization functionality for a broad field of applications beyond LA, there is an increasing number of tools available for the specific purpose of LA, i.e. specifically designed to solve different educational problems (Romero & Ventura, 2013). This refers to tools providing functionality for supporting educational objectives. This includes tools that focus on a particular LA method and/or tools specifically designed for educational application. In the latter case these tools oftentimes center around or add on to existing learning environments – tools of this type are presented separately in the following section 3.2. Table 10 gives an overview of a selection of more stand-alone-type of tools for special LA purposes; two exemplary screenshots are shown in **Figure 17** and Figure 18. This category of tools may focus on different steps of the LA process, like data collection (i.e. supporting gathering or tracking of learning data) or analytics (implementing specific analytics methods) or visualization specifically for LA purposes.

**Table 10:** Special purpose learning analytics tools.

| **Tool name** | **Reference** | **Description** |
| --- | --- | --- |
| BNT-SM (Bayes Net Toolbox for Student Modelling) | http://www.cs.cmu.edu/~listen/BNT-SM/<br><br>Chang, Beck, Mostow & Corbett (2006) | - Matlab package for estimating student knowledge with Bayesian Knowledge Tracing<br>- Extends general purpose Bayes Nets packages for the specific purpose of student modelling |

| | | |
|---|---|---|
| | | - Prediction of student knowledge from his or her observed actions. |
| Degree Compass | http://www.brightspace.com/products/degree-compass/<br><br>Denley (2013) | - tool using predictive analytics for guiding students through course/module selection process<br>- originally developed by Austin Peay State University as course recommendation application for the university; meanwhile part of the D2L eLearning suite[4] |
| myClass | http://next-tell.eu/portfolios/myclass/ | - tool for competence-centered learning analytics and mobile activity tracking<br>- enables teachers to track and record student activities and competence levels to monitor achievement of learning goals and competence acquisition, and to derive activity and competence reports in various formats (e.g. bar charts, competency cloud)<br>- independent from a technical infrastructure and nature of learning activities |
| ProNIFA (Probabilistic Non-Invasive Formative Assessment) | http://next-tell.eu/portfolios/pronifa/<br><br>Kickmeier-Rust & Albert (2013a, b) | - environment for CbKST-related services and functionalities<br>- retrieves performance data (e.g., test results, activities in a virtual environment) and updates the probabilities of the competencies and competence states in a domain<br>- multi- source approach allowing to connect the analysis features to a broad range of sources of evidence.<br>- ProNIFA provides also authoring, analysis, and visualization features. |
| PSLC Datashop | https://pslcdatashop.web.cmu.edu/index.jsp<br><br>Koedinger et al. (2010) | - open data repository of learning data<br>- Provides also analysis and visualization tools through a web-based interface – e.g. displaying student learning over time and patterns of student performance (learning curve analysis) |
| Student Success System | http://www.brightspace.com/products/insights/ | - primarily aims at providing actionable information to teachers by predicting |

[4] http://www.brightspace.com/products/degree-compass/

| | | |
|---|---|---|
| | Essa & Ayad, 2012 | academic progress and identifying students at risks<br>- uses flexible predictive modelling (for application in different educational contexts) and advanced visualisation techniques that provide an overview of factors contributing to success or failure in a course<br>- provides a case management tool for managing and assessing interventions<br>- part of the D2L eLearning suite |

**Figure 17:** Screenshot of myClass (source: http://next-tell.eu/portfolios/myclass/)

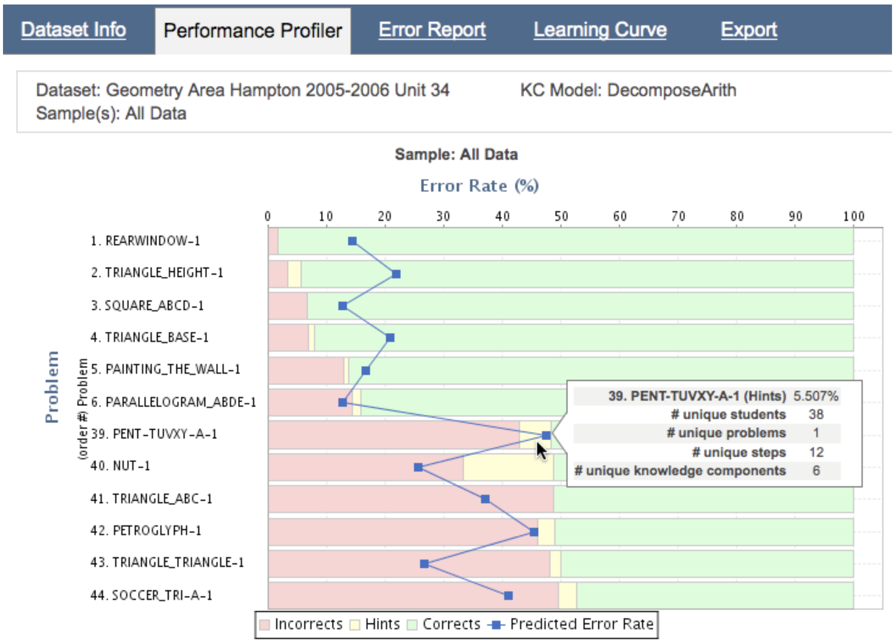


**Figure 18:** Screenshot of the Performance Profiler Tool in PSLC Datashop (adopted from Koedinger et al., 2010).

### 3.2.10 Analytics in E-Learning Systems

LA tools for non-experts optimally should be directly embedded into the standard e-learning environment and tools they are using (Chatti et al., 2012). In fact, many course or learning management systems nowadays provide built-in basic learning analytics tools (Ferguson, 2013). For more sophisticated analysis specific analytics solutions have been developed for educational purposes, which allow for example analyzing learner activity, performance assessment, progress over time on the basis of learning data from an e-learning system and addressing different stakeholder groups and objectives. There are open and commercial tools offering analytics for teaching and learning, which can be used in a 'plug-and-play' manner in combination with existing TEL environments, sometimes even without any input from analysts required. Many providers of learning systems, like IBM or Blackboard, in the context of their e-learning suites meanwhile also provide solutions offering analytics, providing dashboards for administrators, teachers, and learners (van Harmelen & Workman, 2012).

**A very well-known and probably the most cited example of a learning analytics project and tool of this category is 'Course Signals'[5] from Purdue University (Arnold, 2010; Arnold & Pistilli, 2012). It is commonly mentioned in state of the art literature on LA (e.g. Buckingham Shum, 2012; Dyckhoff et al., 2013; Ferguson, 2013; Van Harmelen & Workman, 2012), since it is a relatively early project, which is still still in use. It constitutes a large-scale application of LA and has meanwhile transformed into a commercial product. Signals uses the data collected with an LMS to predict and report individual student student performance, in particular predicting which students are at risk of falling behind or performing**

[5] http://www.itap.purdue.edu/learning/tools/signals/

**badly in the sense of an early warning system. Concretely, the analytics provided by signals categorizes categorizes learners in a low, medium, or high success risk group through colour coding by the use of a a traffic light metaphor (see**

Figure 19). The prediction model and student success algorithm used by Course Signals uses performance data (grades) demographics, as well as information about learner effort (i.e. interaction with the LMS) (Arnold & Pistilli, 2012; Pistilli, Arnold, & Bethune, 2012). The predictions provided by Signals shall enable positive interventions in order to reduce the chance of failing. At Purdue University, Course Signals is a web-based application used in conjunction with the LMS Blackboard; while instructors use Course Signals to check the status of their students and communicate with them, students see their signal directly in Blackboard. The tool is also commercially available for other universities or institutions as part of the Ellucian software suite[6] on student success (Pistilli et al., 2012). Results from an analysis of the use of the Course Signals tool at Purdue have shown positive impact on academic performance - in the experimental group, with the use of the tool, there are higher average grades (increase in A and B grades; decrease in C, D and F grades) and learners visit more help resources than a control group not using the tool; these encouraging results are complemented by positive feedback from learners and teachers (Arnold & Pisitlli, 2012; Pistilli et al., 2012).

A list of other learning analytics tools for application in the context of or in combination with existing learning system is given in Table 11; some exemplary screenshots can be found in Figure 19, Figure 20, and Figure 21.

**Table 11:** Learning analytics tools in the context of existing e-learning systems.

| Tool name | Reference | Description |
|---|---|---|
| Blackboard Analytics for Learn | http://www.blackboard.com/Platforms/Analytics/Products/Blackboard-Analytics-for-Learn.aspx | - analytics application integrated with the Blackboard Learn platform<br>- Dashboards and reports for students, teachers, and administrators<br>- |
| Student Activity Monitor | http://www.role-widgetstore.eu/content/student-activity-monitor<br>Govaerts, Verbert, Klerkx, & Duval (2010) | - tool for visualising time tracking and student activities collected as event in personal learning environments<br>- aimed at supporting reflection and monitoring for learners and teachers; also used as a basis for recommendations of useful material |
| LOCO Analyst | http://jelenajovanovic.net/LOCO-Analyst/index.html | - context-aware learning tool for analytics of learning processes taking place in a web-based learning environment |

[6] http://www.ellucian.com/Software/Ellucian-Student-Success/

| | | |
|---|---|---|
| | Jovanovic (2008) | - aimed at providing teachers with feedback on the relevant aspects of the learning process; feedback is based on analysis of user tracking data |
| Engagement Analytics for Moodle | https://moodle.org/plugins/view.php?plugin=report_engagement<br><br>https://docs.moodle.org/22/en/Engagement_Analytics_Plugin | - plugin for the free e-learning platform Moodle<br>- intended for teachers to provide them with a quick graphical snapshot of students at risk<br>- provides information about student progress against a range of indicators. |
| Compod | http://css-kmi.tugraz.at:8080/compod/web/ | - tool set for competence-based learning<br>- provides visualisations of learning progress for learners (e.g. performance on problems, learning objects visited)<br>- provides detailed course analytics on problems, learning objects, and competences for teachers<br>- includes tools for authoring of domain models, assessment problems and courses<br>- has also been integrated with Moodle |

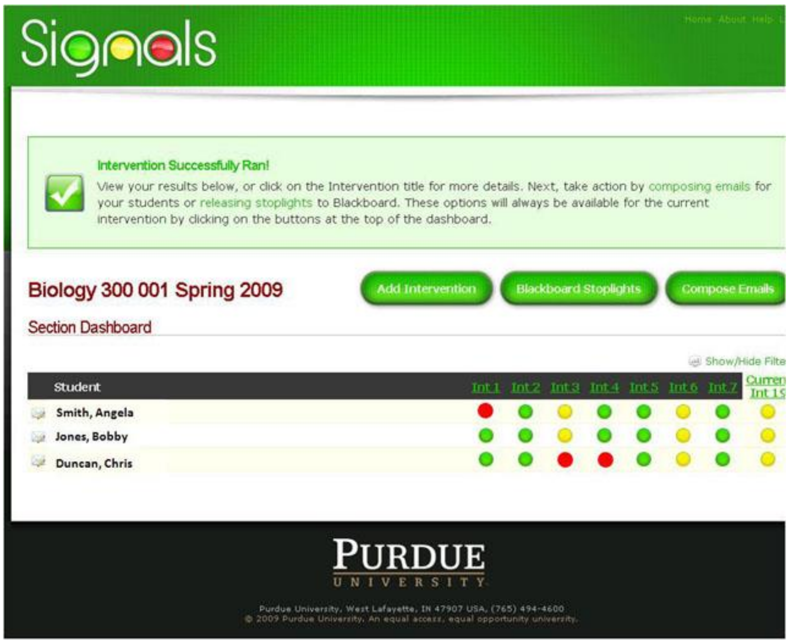


**Figure 19:** Screenshot of the Course Signals interface (figure adapted from Arnold, 2010).

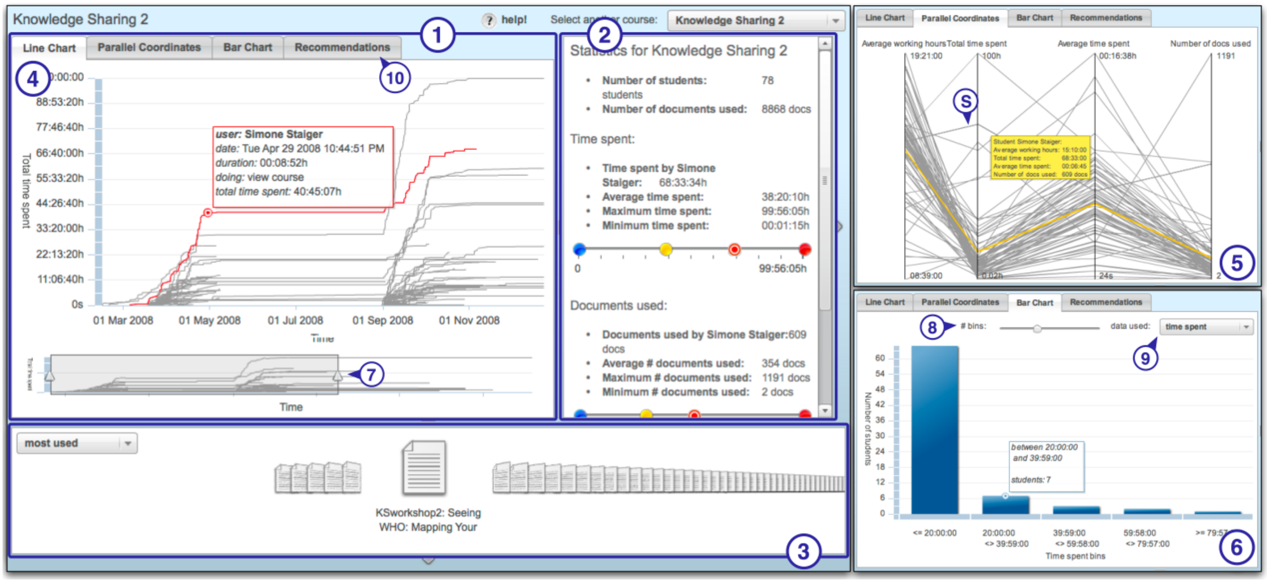


**Figure 20:** Screenshot of Student Activity Monitor
(source: http://www.role-widgetstore.eu/content/student-activity-monitor).

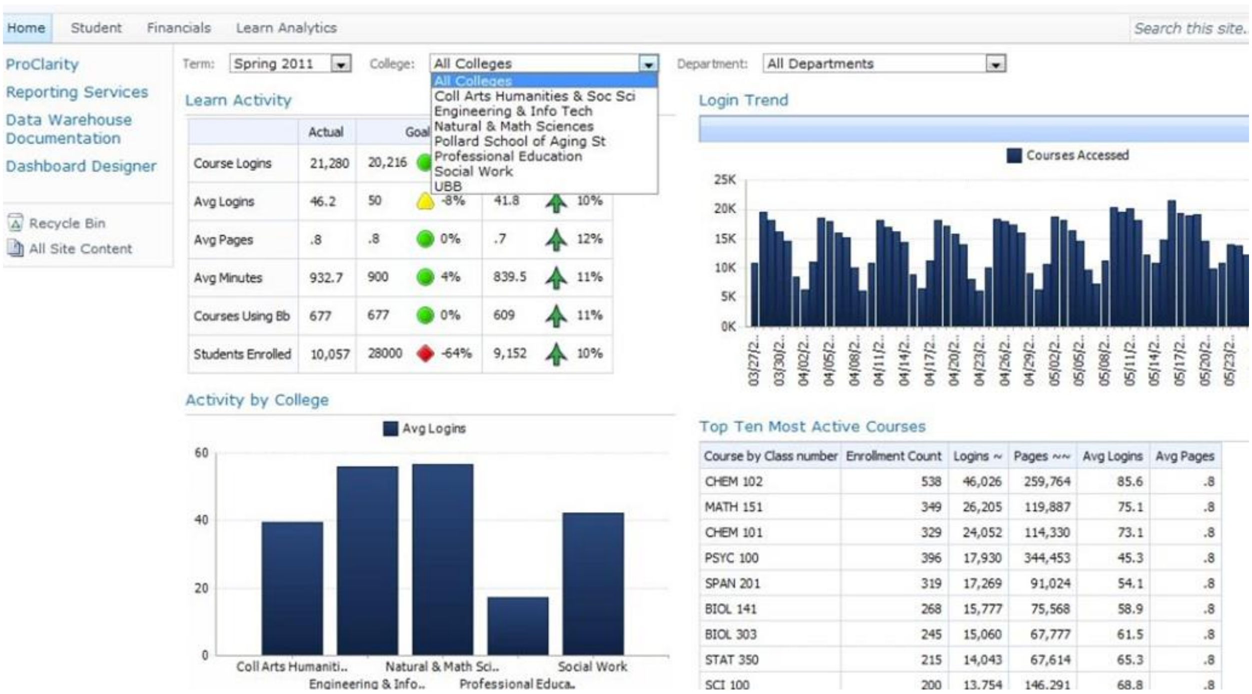


**Figure 21:** Screenshot of dashboard visualization for educational administrators in Blackboard Analytics for Learn
(source: http://www.role-widgetstore.eu/content/student-activity-monitor).

# 4. THE MOST RECENT TRENDS

Over recent years, especially the growing venture capital investments in educational technology companies fueled a slew of technology startups to develop next generation data analytics software to support educational decision making. In USA, venture funding for data analytics tools as of September 2014 is at $58M, which corresponds to a 687% increase as compared year-to-date to 2012. Total funding in 2013 was $75M. These investments are not unwarranted. According to the McKinsey Global Institute, more open data practices in education have the opportunity to unlock an economic value of $800bn to $1.2 trillion annually (this figure includes higher education). Per their analysis, the primary drivers of value include “improved instructional outcomes by identifying more effective pedagogies and cost-savings through more efficient administration.”

Aiming at fully realizing the potential of data analytics these new venture capital backed products are changing the way school leaders, teachers, parents and students think about and utilize data. In the near term, this influx of high-quality data tools will help teachers and administrators answer pressing and challenging questions. Given the sheer amount of paperwork, surveys and test scores generated, education has the potential to be one of, if not the, most data-driven sectors. However, too often the data that teachers and administrators can access are untimely or inactionable. The most recent trends indicate that, this is about to change.

## 4.1 MAKING ACHIEVEMENT DATA MORE ACTIONABLE

The critical questions that K-12 school leaders want answered are rarely captured in a single data set. Data on interconnected issues such as student performance, behavior and attendance are typically housed in dozens of disparate legacy data systems or licensed products that have specific functionality and lack integration.

Products such as BrightBytes and Schoolzilla addresses this problem of multifaceted, multimodal, multidimensional data sources by providing schools with a data warehousing platform that allows educators to connect and clean various data sources into intuitive, actionable visualizations. Once the data is connected, customizable data dashboards enable educators to focus on their key performance indicators. For two example screenshots of Schoolzilla see Figure 22 and Figure 23.

As these products are used by more and more schools educators benchmark their school’s performance relative to other schools laying the groundwork for conversations around best practices. Naturally, some of this conversation is about having better privacy tools and greater levels of integration across systems.

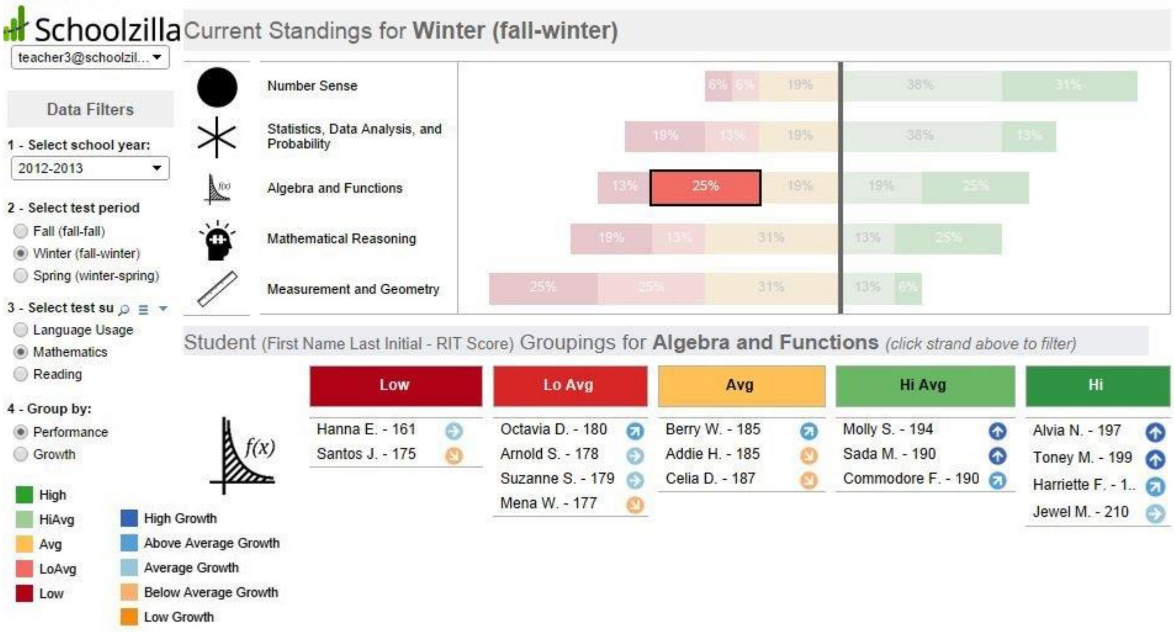


**Figure 22:** Schoolzilla dashboard detailing student proficiency in Algebra and Functions.

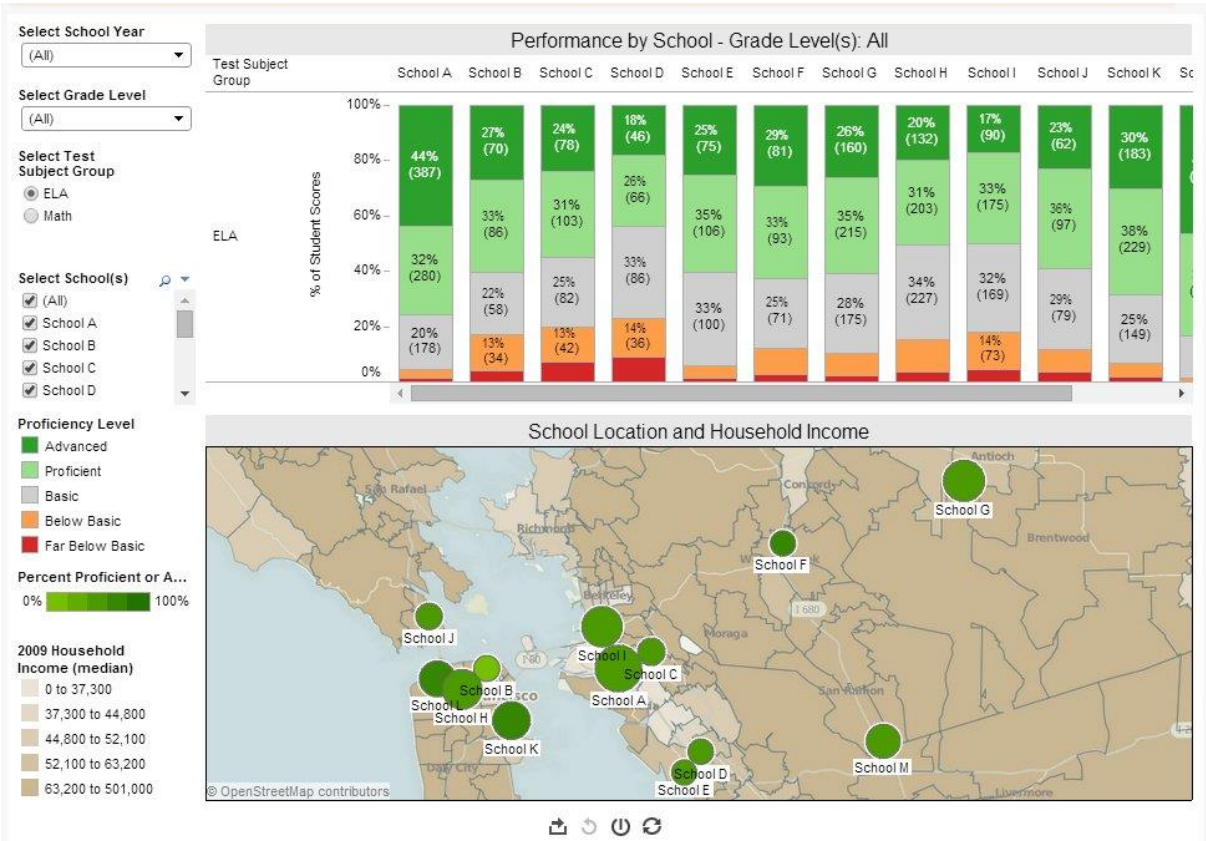


**Figure 23:** Schoolzilla visualizing academic performance together with an interactive map showing average household income by school location.

## 4.2 More holistic portfolio of student performance

Standard tests can only capture a limited component of a developing young learner's academic strengths and weaknesses. Measuring student progress and monitoring achievement gaps have historically been done by such tests, which caused a good part of the educational experience to be oriented towards test success. "Teaching to the test" became an endemic problem, crippling long term and higher order study skills. However, it also presents an opportunity in that replacing evaluation with a more holistic profile of student capability and potential has then the probable consequence of the educational activities to be changed as well.

Freshgrade is one educational collaboration tool that connects teachers, school administrators, students and their parents or guardians. Powered by the data that is accumulated on its network, Freshgrade is aiming to bring a learning-assessment and communication app to K-12 schools far and wide. Another example is Equal Opportunity Schools (EOS), an organization which partners with high school principals and district superintendents to increase enrollment and success of underrepresented students (e.g. low-income and minority). EOS creates data-rich student profiles highlighting academic success and readiness (see Figure 24). Columbia Public Schools in USA have used EOS during 2012-13 school year and doubled the number of low-income and minority students in advanced classes while maintaining or improving exam passage rates for these courses.[7]

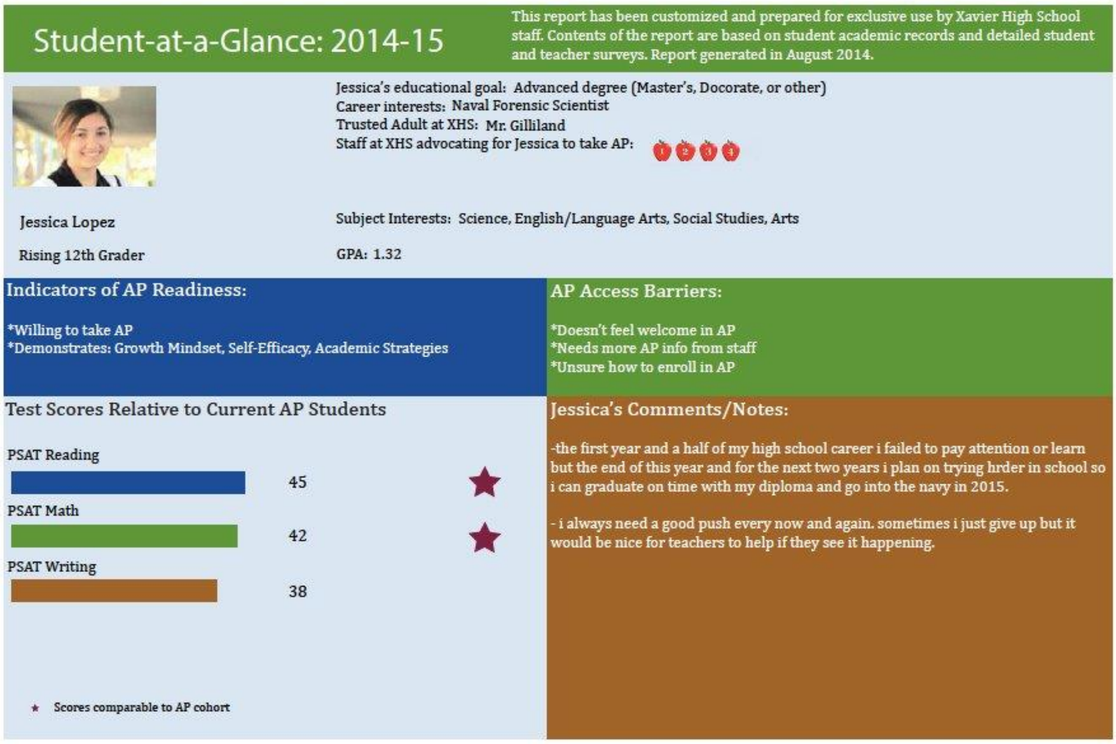


**Figure 24:** A sample student profile created by EOS, which benchmarks performance relative to other students.

[7] http://www.columbiamissourian.com/a/154152/equal-opportunity-schools-helps-recruit-minority-students-for-rigorous-courses/

## 4.3 Giving Students a Voice

Of the new tools that help teachers collect and analyze student feedback, Panorama Education is probably the fastest spreading product. Its student survey reports are elevating the student voice and providing teachers with actionable analytics to understand where there's opportunity to improve instruction and engagement (see Figure 25). This is also a good example of data analytics tools are in impacting at the point of instruction. MasteryConnect, which raised $15.2M funding last Summer also runs a very similar business plan. It leverages an educational collaboration network for assessment and lesson planning to generate analytics based student reports (see Figure 26).

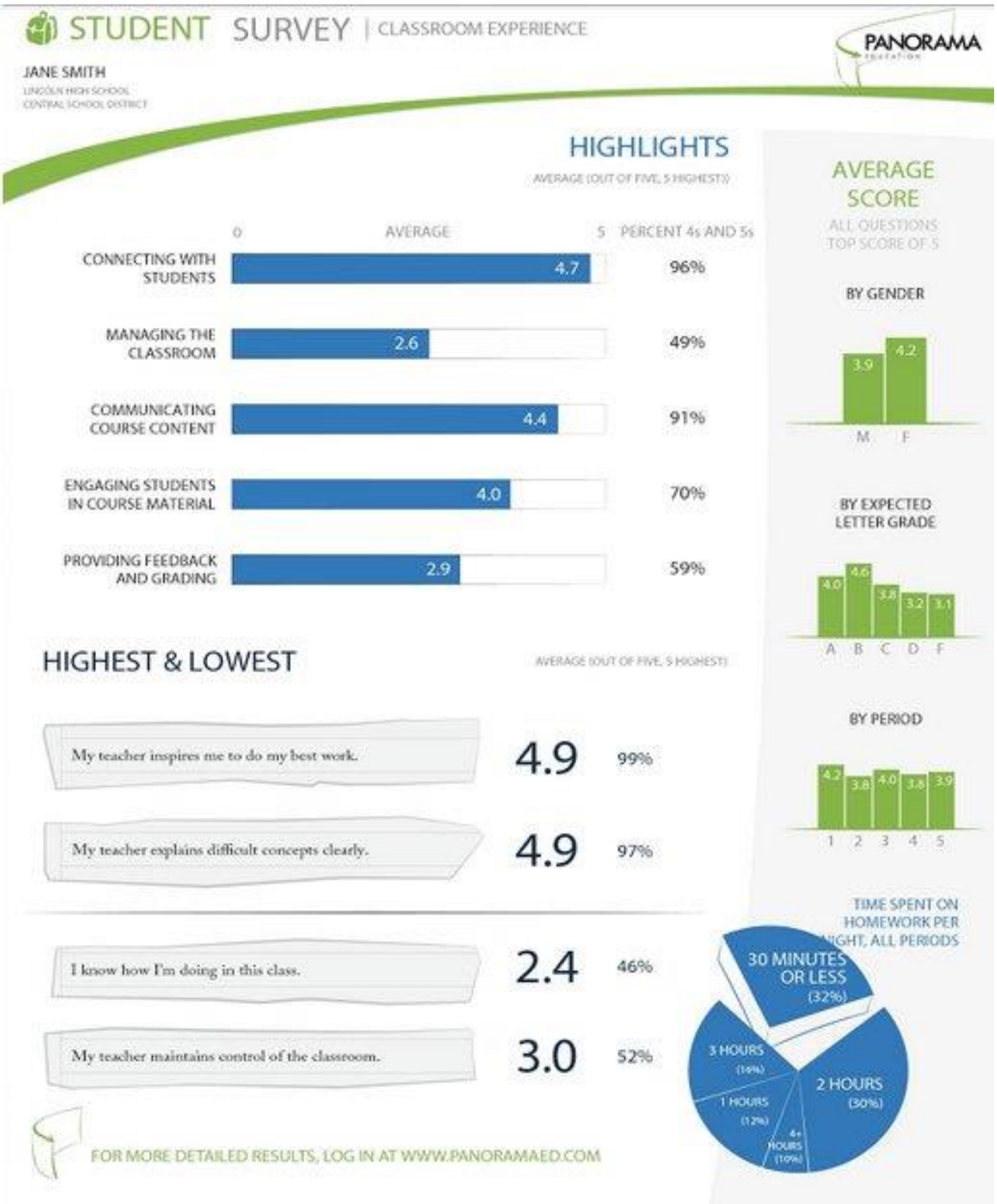


**Figure 25:** A sample Panorama Education report used by teachers.

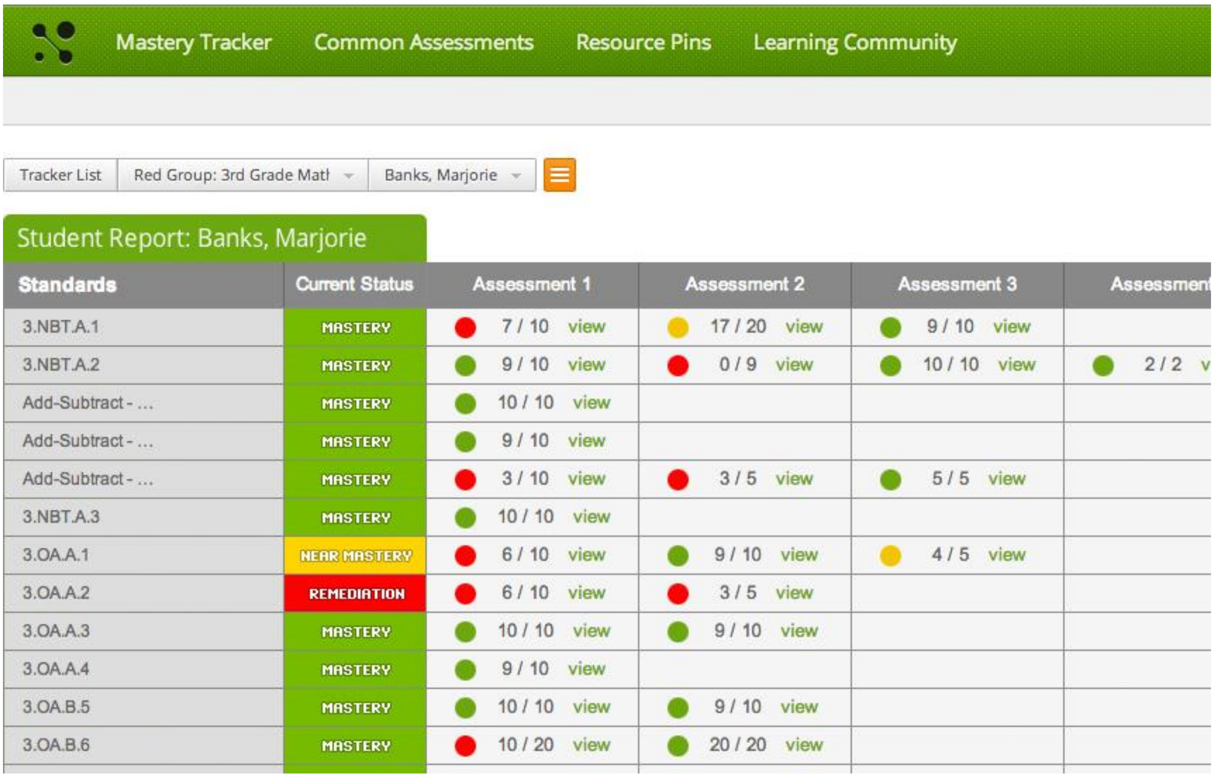


**Figure 26:** Mastery Tracker sample screen from MasteryConnect.

## 4.4 Adaptive Learning

Among the learning analytics companies Knewton attracted the largest amount of investment in the world. Receiving $50M in 2013, Knewton endorsed its adaptive learning platform in a number of countries in the world. Using performance and interaction data Knewton utilizes a set of crunching algorithms to adapt learning material to students in the form of recommendations, in order to minimize their class preparation time.

Ellevation Education is aiming at achieving similar success in adaptive learning for English language learners. While all students have unique needs, ELLs can pose certain complexities rooted in cultural and linguistic differences. Ellevation provides reports on student’s strengths and challenges to help educators personalize instruction for English Learners (see Figure 27).

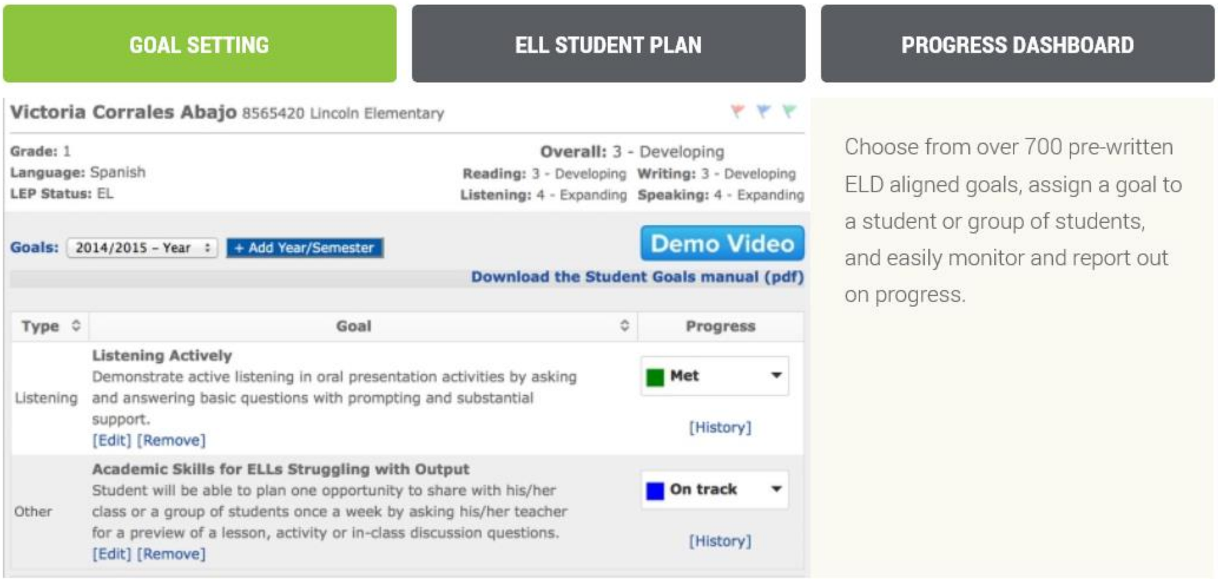


**Figure 27:** Ellevation Progress Dashboard.

## 4.5 Institutional Analytics

Virtually every school runs a student information system (SIS), which in some countries is even linked to a nationwide datamart. SIS is where student records are created, school schedules are organized, grades and attendance are stored, and much more. These systems are very good at ingesting lots and lots of data but they are notoriously difficult to pull data out of. BrightBytes and LearnSprout are among the new learning analytics products that use SIS data to make better decisions about improving operations at a school-wide or even nation-wide level.

LearnSprout produces administrative reports, among others on student health (based on attendance) and college readiness (based on test scores). Brightbytes which received $15M venture funding this year, is used by 10,000 schools, aiming at evaluation and improvement of technology use in the classrooms. Class level analytics is not only supports administrative decision making, but also help direct teachers' professional development.

# 5. Excursus: LA in Virtual Worlds and Serious Games

Digital learning games are another development increasingly recognized by educational practitioners as useful educational tools with highly motivating character (e.g. Johnson et al., 2014). With the application of virtual worlds and games to support instruction and learning, there comes a need of acknowledging learning game experiences also in the context of educational assessment. LA is considered as a promising approach for supporting performance measurement, assessment, and improvement in and of serious games and virtual worlds. Thus, game-based learning is considered a research area and technology that should be explored from a LA viewpoint (Papamitsiou & Economides, 2014). Initial steps of translating and using LA also to these types of educational environments have already been taken.

Virtual worlds are highly engaging, offering opportunities for learning experiences that go beyond traditional e-learning environments (Camilleri, de Freitas, Montebello, & McDonagh-Smith, 2013). With higher education institutions and universities starting to offer their courses in such online environments and enterprises specializing in the delivery of experiential corporate training, educational administrators and training providers are also starting to use analytics in order to get to know about the number of learners signing up these courses and how they are engaging with the course material and with each other. Learning analytics is considered as a way of better understanding the learning pathways of learners in virtual worlds, in order to identify the effectiveness of this kind of training, to foster reflection on the learning and teaching process, to modify teaching approaches etc. Attempts of applying analytics in virtual worlds have been presented e.g. by Camilleri et al. (2013), Fernández-Gallego, Lama, Vidal, and Mucientes (2013), and Kickmeier-Rust and Albert (2013a,b). Fernández-Gallego et al. (2013) propose a process mining-based learning analytics framework in order to capture the actual learning pathway taken by learners in a virtual world, and thus enabling teachers to understand which additional learning activities not initially foreseen in their learning design have been performed and to evaluate the efficiency of their course and whether the new activities contributed to the educational objectives. Camilleri et al. (2013) mainly aimed at assessing the level and depth learners' engagement with the learning task and their interaction with objects and other learners in a virtual world. The analytics uses a dashboard visualization and mainly serves trainers and administrators to evaluate the benefits of trainings provided and monitoring user activities, for example to identify learning tasks or objects particularly effective for promoting learning. Kickmeier-Rust and Albert (2013a,b) describe a competence and learning performance-oriented log-file analysis and visualisations of learning activities in virtual worlds, which may enable teachers in making more effective use of this kind of educational scenarios.

While teaching scenarios in virtual worlds resemble more traditional teaching environments and provide some formal learning context, in educational games learning is usually embedded in the context of a game, ideally realizing a stealth learning environment. There is a broad awareness and agreement on the educational potential of videogames and game-based learning and serious games are intensively researched (e.g. de Freitas, 2013). Games are highly motivating and appealing; the interactive and immersive learning experiences that can be created by the use of learning games establish authentic learning tasks and meaningful, situated learning. Games have proven to support skill acquisition related to collaboration, procedural and critical thinking, creative problem-solving, observation, reasoning, and collaboration (Johnson et al., 2013).

Despite the theoretical and empirical evidence for the potential of educational games, there is still some reluctance among teachers towards their broader take up and use in educational practice

(Serrano-Laguna, Torrente, Moreno-Ger, & Fernández-Manjón, 2014). This is mainly due to the fact that it is difficult to integrate assessment procedures in terms of tests or question and answers in games, which would be perceived highly disruptive and breaking the flow of the gaming experience (Van Eck, 2006). Assessment routines built in commercial games, even if they are developed for educational purposes, are usually black boxes and not tangible for teachers. Serrano-Laguna, Torrente, Moreno-Ger, and Fernández-Manjón (2012a) highlight that there is a need to implement approaches for reliable formative and summative assessment in educational games, which are easy to use and provide teachers useful educational information and evaluation. Learning games, just as commercial video games, may produce large amounts of data by recording user (inter)actions on a micro level. This results in another type of learning-related big data that may be used for LA. A crucial question is how to harness and make sense of this data in an effective and efficient manner. The LA community is currently in the process of initiating the elaboration of analytics that can be used on serious games. By using and combining ideas from gaming analytics, web analytics and learning analytics for (serious) game data it is possible to establish meaningful analytics for educational purposes. LA in serious games may serve two main purposes, which are mostly in line with general objectives of LA (as presented in section 2.4): 1. Measuring student success for feedback to learners and teachers, 2. dynamic adaptation during gaming. In addition LA may also be exploited to refine or improve educational games – in terms of using analytics for an evaluation of the game artefact itself (Serrano-Laguna et al., 2012a). LA for serious games may be implemented by following a generic approach or by designing LA into the game. The latter one involves more effort but is considered to have potential of providing deeper understanding of learning.

A great challenge with LA in educational games is the wide variety of different games available, which complicates the development of analytics tools that are applicable independent of a concrete game. To overcome this, Serrano-Laguna et al. (2012a, 2014) propose a two-step generic approach to apply learning analytics in educational games, which is applicable with any kind of different game. First, generic traces are gathered from gameplay, including game traces (start, end, and quit), phase changes (game chapters), input traces like mouse movements or clicks, and other meaningful variables like attempts or scores (depending on the game). These data give rise to reports with general and game-agnostic information, like the number of students who played the game, average playing time, game phases in which users stopped playing etc. This information can be visually reported and may provide initial useful information on how learners interacted with a game. In a second step, additional information may be extracted by letting teachers define game-specific assessment rules based on and combining the generic game trace variables to obtain new information (e.g. setting maximum time thresholds, comparison between actual and expected/required values of variables). These rules clearly need to be closely defined in line with each game to match the educational objectives; however since the building blocks of this kind of rules are elements from the basic set of traces the creation and provision of template rules to support teachers in defining their own are conceivable.

To make use of learning analytics in educational games, necessarily a game platform needs to be used that allows collection of the relevant data, and that holds a representation of game variables. The data for learning analytics will likely need to be stored and processed separately and remotely To technically implement such kind of analytics in an educational game, the definition of a learning analytics model and implementation of a learning analytics engine, which is separate from the game engine but communicates with it, has been proposed (Serrano-Laguna, Marchiori, del Blanco, Torrente, & Fernandez-Manjon, 2012b). The learning analytics engine is conceived as comprising a set of modules enabling the different steps of the learning analytics process, from capturing data, via aggregating and reporting to evaluating in terms of transforming information in educational knowledge.

Apart from a generic approach to LA in serious games, analytics may be designed and implemented into the game for assessment and monitoring. Assessment in a learning game may thereby have two main purposes: Firstly, just measuring the success of the student – this will serve providing teachers and students the information derived as a basis for action, like selection of new educational resources, decision on additional support or learning tasks etc. Secondly, the derived information may be used for realizing dynamic adaptation during game time through an adaptation model and adapter (part of the learning analytics framework) communicating with the game engine. In well-designed educational and serious games learning the great challenge is to align story and lesson, such that learning occurs embedded in the storyline of the game (Shute, Ventura, Bauer, & Zapata-Rivera, 2009). In line with this, and in order to avoid disrupting game experience, assessment also needs to be embedded in the game. Shute uses and evidence centered design approach as a basis for implementing so-called stealth assessment in educational games. This approach starts from the definition of educationally relevant competencies, defines evidence models linking these competencies to game behaviour, and regularly updates the user model on learner competencies through Bayesian models. In this way, learning can be monitored and fostered, for example by generating progress reports and selecting new game experiences (Shute et al., 2009). In addition, Ventura, Shute & Small (2014) elaborated a method for assessing persistence in educational games based on the time that learners spent on unsolved problems, which may be used to tune gameplay difficulty, feedback and hints. Continuous assessment and adaptation has been elaborated and implemented in the educational game demonstrators developed in the ELEKTRA and 80Days projects (Kickmeier-Rust & Albert, 2010). In a nutshell, learner actions during a complex problem-solving situation are monitored and interpreted in run-time to enable a continuous and non-invasive assessment of learning progress and motivational state. Observations of learner actions are interpreted as evidence for available and lacking skills and feed a probabilistic assessment and continuous update of the learner's competence state (Augustin, Hockemeyer, Kickmeier-Rust & Albert, 2011). The information coming from this assessment is used to provide adaptive interventions tailored to the individual's current state and needs (chosen from a menu of different intervention types), to support and guide the learner in the game and learning task and to retain motivation (Kickmeier-Rust, Steiner, & Albert, 2011). The analytics applied in these educational games were dedicated to realizing continuous assessment and automatic live adaptation and support. This application could be broadened in terms of reporting and feeding back the information on skills acquired and learning progress also to learners and teachers, thus leveraging the educational value of the analytics processes of the games and translating the learning data from the game into educational actions outside the game context. Another example of using learning analytics in a serious game has been presented by (Baker, Habgood, Ainsworth, & Corbett, 2007), who also realized skill assessment in an educational action game by using game events as evidence for users' mathematical skills to analyse study gains in accuracy over time and in speed over time with learning curves. This approach proved useful for formative assessment in educational games and may also be used to inform re-design and improvement of intelligent tutoring systems. Another very recent LA attempt has been made towards elaborating an automated detector of engaged behaviour in a simulation game (Stephenson, Baker, & Corrigan, 2014). The aim thereby was to identify and model which learner actions give evidence of user engagement and, in the end, are predictive for success in the game. An integration of the engagement detector in the game will enable to report the results back to learners and teachers for reflection.

On the whole, LA has started to grow into the field of serious games, but there is still much more work to do to fully exploit the potential that LA may bring to optimize learning experiences with educational games. In particular, it needs to be taken into account that games may be part of multiple learning activities and tools that learners carry out and use in parallel and, potentially, on the same educational objective or domain (Miller, Baker, & Rossy, 2014). A learning game will usually provide educational content as a complement to other and more traditional technology-enhanced and classroom learning activities. LA for assessment of skill acquisition should therefore actually not consider an educational game in isolation, but should optimally utilize and integrate learning game data with learning data from other sources for a holistic understanding of learning.

# 6. Drawbacks and Challenges of Current LA Approaches

Although a recent trend of moving from a technological focus towards a more educational direction can be identified (Ferguson, 2012), Siemens (2012) highlights that there is still a '*research and practice gap*' that exists in learning analytics in terms of a lack of translating analytics research into educational practice: "The work of researchers often sits in isolation from that of vendors and of end users or practitioners" (p. 5). The focus of LA research until now has been on the methods of data collection and analysis instead of the actual application in educational practice. A thorough elaboration and testing of LA methods before deploying them in practice is, in fact, meaningful. But with LA methods getting more mature, a transition in analytics from technical orientation to one that emphasizes sense-making, decision-making, and action is required to increase interest among educators and administrators. A second transition needed in LA: one that moves LA research and implementation from at-risk identification (which is only a small aspect of what analytics can do to improve education) to an emphasis on learning success and optimization.

In addition, to ensure practical significance and usefulness of LA approaches, the consideration and, optimally, involvement of stakeholders' point of view is important. Teachers show high interest in evaluating their courses and teaching (in the sense of action research) and have a range specific research questions in mind (for an overview see Dyckhoff, 2011), but at the moment there a lack of appropriate LA tools enabling to systematically and efficiently investigating those questions. This becomes evident from a meta-analysis identifying the kind of data and indicators used in state of the art LA and determining which kind of research questions from teachers these are able to answer (Dyckhoff et al., 2013). Analytics sometimes use still very basic measures and indicators, while lacking a deeper consideration of how to translate the educational data into meaningful information on learning, knowledge acquisition, or experience of learners. The power of LA approaches used today to measure complex processes, such as learning, is oftentimes limited. It has been pointed out that instead of data driven analytics, LA needs to ground analytics more in the learning sciences in order to deal with the complexities of learning processes and establish a holistic picture of student progress and lifelong learning (Ferguson, 2012). Inspired from action research, a method for professional development and improvement of teaching, LA should start from research questions that arise from teaching practice, as opposed to the traditional approach of starting analytics based on already collected and available data (Dyckhoff et al., 2013). Even more, instead of providing answers to questions that may be answered also in different ways, LA needs to pursue providing new perspectives and additional insight on learning data. Building a strong connection with learning sciences constitutes a challenge for LA research, but at the same time provides opportunities of more theory-grounded and educationally relevant LA approaches. With the adoption of increasingly competence-oriented teaching practices (e.g. Müller, Gartmeier, & Prenzel, 2013) and the importance of formative assessment in education, there comes also the need of advancing LA approaches providing an insight on skill acquisition and competence development (Kickmeier-Rust & Albert, 2013). Analytics, besides using pure reports on interaction data, should support deeper analysis in terms of the assessment of knowledge or skills. In fact, such kind of knowledge- or competence-based interpretation appears necessary in order to provide meaningful information for further educational action and learner support. Besides, taking into account aspects like sentiment, affect, or motivation in LA, for example by exploiting novel multimodal approaches (e.g. Bahreini et al., 2014), may provide a deeper understanding of learning experiences and the possibility to provide educational interventions in emotionally supportive ways. Besides, integrating LA in daily educational practice should be accompanied by suitable didactical and organisational frameworks (Chatti et al., 2012). Advancing LA in these directions will foster to educational relevance, acceptance of and interest in of analytics approaches.

With respect to analytics techniques and tools available one problem is that findings from marketed analytics software are most often proprietary and therefore not available to other researchers, which in turn hinders quick and iterative improvement cycles of learning analytics methodologies. Corporate analytics products are closed to researchers and do not allow to access and scrutinize, change and improve algorithms. There are only few tools providing the openness, accessibility, and transparency desired by researchers (one example is the software package R) (Siemens, 2012). LA tools are sometimes still difficult to understand, not well organised, and often provide information only in tabular format (Dyckhoff, 2011), but increasingly visual approaches of reporting in form of dashboards can be identified (Buckingham Shum, 2012). LA research sometimes seems to tend to develop more and more complex analytics systems to gather learning data for later analysis (Mohamad & Tasir, 2013). Usability and completeness of results are common issues, as well as the lack of integration of different data or results (e.g. from questionnaires and from interaction logs) (Dyckhoff, 2011). This is especially relevant when considering educational application of LA. Many existing tools focus on application for research purposes, they are very complex and their features go far beyond the scope what a teacher may need. On the other hand, many questions relevant for teachers in educational practice cannot be answered with analytics tools that are currently available (Dyckhoff, 2011). For application in educational practice, more tools are needed that are suitable for teachers and non-expert users – tools with intuitive user interfaces and good visualisations, tools that are easy to use and provide features that are meaningful to educators and allow positive end-user experiences (Romero & Ventura, 2010; Siemens, 2012). In the design and implementation of LA tools for actual educational deployment, the involvement of and engagement with teachers and learners is therefore necessary to ensure that the tools developed actually address relevant educational questions and meet the intended objectives of improving teaching and learning (Chatti et al., 2012).

Another challenge in LA is given by the fact that the amount and diversity of educational data is ever growing, which makes the analysis of educational data increasingly complex and increases demands on data storage and computational power. This data explosion in the educational sector itself impacts analytics – with increased quantity of data the methods and approaches used for analysing and making sense of this data need to change, too (Long & Siemens, 2011).

Even more important, though, is the need for *integration and interoperability* of learning data from different sources. When learners are dealing with online learning environments and tools, they will most likely not be engaged with on single learning activity, but will instead carry out various different activities or learning tasks at a time (Miller, Baker, & Rossi, 2014). LA, for example approaches of assessing learners' knowledge or skills, is nowadays usually still confined to only one activity (and data source), also if the same skills are involved in different parallel learning activities. An increasing awareness is developing for this need of integrating learning data from different sources to build more comprehensive and conclusive learner models and to derive more targeted conclusions for supporting or optimizing teaching and learning. Existing approaches in this line of thought are still limited; for instance they concentrate on predicting learner performance on one activity on the basis of data from another activity or, respectively, investigating whether learning transfers to new contexts. Miller et al. (2014), for example, integrated information from conceptual instructions, problem-solving, and mini games for predicting student performance using Bayesian Knowledge Tracing. While this work used learning data on a small set of different activities, but stemming from one and the same learning system, there have also been initial attempts of tracking meta-skills of science inquiry across different knowledge domains but within the same activity (Sao Pedro, Baker, Gobert, Montalvo, & Nakama, 2013).

To summarise, “one big problem around learning analytics is the lack of clarity about what exactly should be measured to get a deeper understanding of how learning is taking place” (Duval, 2011, p. 16). Currently used indicators mostly are limited to interaction data (Dyckhoff et al., 2013), but how much information can these actually provide about the learning process? LA results sometimes still consist in very basic measures and indicators, while lacking a deeper consideration of how to translate the educational data in meaningful information and to measure the complex processes of knowledge and skill state or acquisition of learners (Siemens, 2012). LA needs to go beyond tracking and reporting basic usage data to make inferences on the knowledge and competence of learners, their affective states etc., and it needs to include more than data from only one centralized learning system (Johnson et al., 2012), but should include also data from different tools and from more informal learning, to establish an accurate and deep understanding of learning and teaching. In addition, teacher data is usually not considered and explicitly collected in current LA. There is a need to include indicators on teaching and teaching activities in LA, and to relate them to student learning (Dyckhoff et al., 2013). To ensure that LA develops towards answering the complex research questions that are relevant for practitioners, emphasis needs to be put on the elaboration and use of such high-level and combined indicators, and educators should optimally be actively involved in the definition and design of relevant indicators.

An important issue that needs to be taken into account is the *possibility of error*. LA is only as good as the data underlying it. If an analytics approach has to rely on data hat is not or not accurately or sufficiently tracked, the results from analytics will necessarily be less than optimal and steps need to be taken to record or improve the data (van Harmelen & Workman, 2012). In addition, although LA optimally shall pursue a continuous process of improvement, the results and predictions derived from LA are not necessarily perfect or valid, and usually not all possible data sources or causes/indicators for success or performance will have been taken into account. So the question is what are the consequences of such errors, and of interventions or decision taken based on them (Campbell & Oblinger, 2007). Deploying analytical tools that do not work totally reliable or without flaws may compromise confidence in analytics and may create tension and frustration in an educational institution (Crow, 2012). It has therefore been underlined that it is important to allow overrides for decisions, which are taken from learning analytics (Davenport and Harris, 2007, cited after van Harmelen & Workman, 2012).

Finally, a great challenge in LA relates to *ethical and privacy issues*. Ethical and privacy issues can be considered as orthogonal to the different steps of the LA process. These topics have been dealt with some tension in LA (Pardo, 2014). There is a trend of considering users as the owners of the data collected about them and institutions are borrowing them for a clearly stated purpose. In LA things get more complicated very quickly, since usually data from a whole population of learners is used to produce a prediction model – and the question arises who the owner of such kind of model is (Pardo, 2014). At the moment there are no standard methods and procedures for informed consent, opting out etc. There is a need to develop a clear and agreed set of ethical guidelines with respect to the ownership of data and analytic models, rights and responsibilities (Ferguson, 2012). In any case, choice and consent from learners need to be inherent part of an LA approach and transparency of what data is used and how it is analysed should be given to users. Principles to be taken into account by educational institutions are (Slade & Prinsloo, 2013): LA is a moral practice and needs to focus on understanding instead of measuring, learners are central agents and collaborators, learner identity and performance are dynamic variables, learning success and performance is complex and multidimensional, data collection and processing needs to be done with total transparency, and education cannot afford not to use data. It has been suggested that to approach and resolve ethics and privacy issues in LA one should have a look who these issues have been addressed in related areas, like business analytics, to see how solutions used there can be translated to LA (Pardo, 2014). The urgent need for a uniform approach or framework for dealing with ethics and privacy in LA is increasingly acknowledged and attempts are being made for advancing LA in this direction (e.g. Slade & Prinsloo, 2013; Willis, Campbell, & Pistilli, 2013; Willis, 2014). Kay, Korn, and Oppenheim (2012) outline that LA is in the area of conflict between assuring educational benefits, business interests of educational institutions, and expectations of the born digital and born social generations and postulate four key principles for good practice with respect to ethical aspects and analytics above the level of these conflicts:

- **Clarity**: definition of purpose, scope and boundaries
- **Comfort** and care: consideration of interests and feelings of the data subject
- **Choice and consent**: information and opportunity to opt-out or opt-in
- **Consequence and complaint**: acknowledging the possibility of unforeseen consequences and mechanisms for complaint

These principles and the work on ethical issues and approaches in LA and big data, in general, provide a sound basis for guiding researchers and institutions to address ethical issues in appropriate ways.

# 7. EMPIRICAL EVIDENCE OF LA

Learning analytics sometimes are still on a rudimentary pedagogical level (Shum, 2011) and the need for establishing links to learning sciences and educational practice has been discussed in the previous section. A further need is impact analysis: in addition to focusing on supporting teachers and learners in their tasks, the effect of LA on stakeholders' behaviours and teaching/learning outcomes needs to be proven. LA needs to offer information and support that is useful and not accessible in current educational practice and needs to provide measurable benefits for teaching and learning. Only then educational organisations and decision makers, teachers and learners will be willing to adopt and use these tools. By providing empirical evidence and measurable impact of LA, reservations regarding ICT use, in general, and LA technologies, in particular, can be overcome and an awareness and understanding for the potential of this field of research and development as a means to shape the future of education can be developed.

Papamitsiou and Economides (2014) have conducted a systematic qualitative review of empirical research on LA from 2008 to 2013. Thereby, 'empirical research', in general, was taken into account. The review thus included papers reporting on different kinds empirical studies and results: Studies with more algorithmic, methodological research objectives investigated, for example, the predictive value of different types of indicators, or were dedicated to validating or optimising analytic models (e.g. Huang & Fang, 2013. Papers included in the review that reported more pedagogy-oriented findings included studies that used LA methods for investigating learning/teaching behaviour, like finding relationships between behavioural patterns (e.g. participation) with performance (e.g. grade) (e.g. Giesbers, Rienties, Tempelaar, & Gijselaers, 2013). Only a small subset of the studies considered in the review of Papamitsiou and Economides (2014) actually targeted the educational impact of LA, i.e. the effect on and added value for teaching and learning (Barla, Bieliková, Ezzeddinne, Kramár, Simko, & Vozár, 2010; Lin, Hsieh, & Chuang, 2009)

In general, there is a lack of studies providing empirical evidence of the impact of learning analytics, i.e. the effect of LA on learners' and teachers' behavioural reactions. There have been only few large-scale deployments and applications of LA. One popular example is Course Signals (see section 3.2.10), which has been widely applied at Purdue University and for which positive effects on academic performance and user feedback could be shown (Arnold & Pistilli, 2012; Pistilli et al., 2012). With Course Signals, teachers are provided with information on which students are at risk of railing or dropping out and supported in educational interventions. An example for the positive impact on learning outcomes in the context of a system using LA for automatically adapting instruction to leaners are the Cognitive Tutors at Carnegie Mellon University. Such tutors are available for different knowledge domains, the most widely distributed is one for algebra, for which better end-of-course learning performance could be empirically proven (Koedinger & Corbett, 2006). The evaluation studies conducted on LA to date are oftentimes limited to small-scale laboratory or field investigations and providing limited generalizability; although suggesting benefits for learning, many of these approaches have not been applied and validated at large scale (Baker, 2014; Dyckhoff et al, 2013). Verbert et al. (2013) reviewed fifteen LA technologies using dashboard approaches from which two thirds have been empirically evaluated. Apart from a large scale and longer-term study on the above-mentioned Course Signals (Arnold & Pistilli, 2012), all the other studies were limited, in terms of small samples and short-term use of the technologies usually highly controlled study settings. Most importantly, evaluation studies oftentimes assess constructs like perceived usefulness or satisfaction, or even focus on functionality and usability issues, while failing to demonstrate the beneficial impact of the LA technology on learning, like change of behaviour, improvement of learning performance etc. (Dyckhoff et al., 2013; Verbert et al., 2013). Users' acceptance of LA tools is an important aspect when considering real-world application of LA, since it is associated with users' actual intention to use those tools. The consideration of user acceptance parameters, like perceived usefulness and perceived ease of use, in empirical studies is therefore reasonable. Aside from traditional models on technology acceptance (Davis, Bagozzi, & Warshaw, 1989) or computer-based assessment (Terzis & Economides, 2011), also specific approaches of modeling and assessing acceptance in the specific context of LA have been devised (Ali, Asadi, Gasevic, Jovanovic, & Hatala, 2013).

LA tools should, in fact, not only be usable and interesting to use, but also useful in the context of the goals of LA. Dyckhoff et al. (2013) have proposed an impact evaluation method for measuring this impact. Such evaluation study needs to be conducted with a representable sample of (non-expert) teachers and learners. In a first step a clear picture about the current state of the classroom situation should be gathered through qualitative and quantitative methods and with teachers and learners, in order to have a reference point for later comparison. Based on the information about the current situation, hypotheses can be formulated how the use of the LA tool is expected to affect learners' and teachers' behavior and activities. Then LA tools should be made available to users and, after an appropriate time, to collect again evaluation data. Comparing the results from questionnaires/interviews, performance, log data to those collected at the beginning of the evaluation, changes in the learning and teaching process can be identified. Comparing these actual changes to the hypothesized changed, it is possible to draw conclusions on how the expected impact of the LA tool relates to its actual impact in empirical application. This evaluation approach is also considered to provide a lot of useful information for iteratively refining existing and informing design of future LA tools (Dyckhoff et al., 2013).

When applied in instructional practice, analytics should aim at being conducted in (near) real-time, learners will receive notification of conceptual errors earlier than they currently do when the educators marks exams or essays. For educators, real-time analytics and visualizations will identify challenges facing different learners based on concept comprehension or through sentiment analysis of discourse. Analytics provides an institution the opportunity of moving from simple understanding and reporting of various data points to using them for generating actionable intelligence and taking action on that intelligence, to positively affect educational behaviours and outcomes. The impact or usefulness of LA for learners (e.g. helping learners improving their learning and performance), teachers (e.g. providing the possibility of rapid changes in pedagogical practice), and educational institutions (e.g. improving graduation rates) has been conceptually elaborated in Pistilli et al. (2014), however, without providing any empirical results support these.

At the moment there is little evidence available in the literature elaborating on how learning analytics influence practical educational situations and the behavioural reactions of teachers and learners. The consideration of how learning analytics affects teaching and the evaluation of this impact have been highlighted as crucial issues that need to be tackled in learning analytics research in the future (Dyckhoff et al., 2013). LA tools need to be deployed in educational practice in continuous exchange between LA researchers and teachers learners, to enable a continual improvement and advancement of LA strategies and methods in line with these stakeholders' actual needs.

Drachlser and Greller (2012) did an evaluation study on learning analytics with a somewhat different perspective: they tried to evaluate the understanding, perceptions, and expectations of educational practitioners and researchers in learning analytics. One main conclusion from their survey was that there is, in fact, still low awareness of LA among stakeholders. This issue, however, is assumed to reduce over time, with the rise and spreading of useful LA applications.

At the moment, there are still far more examples of LA technologies published than actually applied in real-world teaching scenarios as a basis for driving decision making and interventions (Baker, 2014). Different educational institutions will be more or less ready to adopt and introduce analytics for teaching and learning. First and foremost, a necessary requirement is the availability of the required technical infrastructure in schools (Wastiau et al., 2013). Moreover, digital and ICT skills need to be available or trained in educational institutions (e.g. Ferrari et al., 2014) in order to enable the successful application of ICT in day-to-day teaching and learning, and thus, the logging and accumulation of learning-related data needed for LA. Two main risks of adopting analytics are that analytics may provide measures that are actually not useful in educational practice and educational practitioners and learners may lack appropriate knowledge how to use analytics (van Harmelen & Workman, 2012). Importantly, it is not sufficient to provide educational institutions access to learning analytics tools, and to provide access to data and analytics results. To have an actual benefit and impact for the education sector, institutions need to train their staff in using, contextualising, interpreting, and acting upon the obtained data (MacNeill et al., 2014). With LA there might be a risk of focusing attention on the consideration and achievement of a limited set of measurable criteria, while disregarding other relevant aspects of learning. Educators therefor need to take into account and be aware of the methods and teaching and learning models underpinning analytics, and the variables and categories of events or (intera)action used and prioritised by these models (Ferguson, 2013).

To summarise, LA has been recognized as an emerging technology that is assumed to have a large impact on the practice of education, with a time-to-adoption horizon of one to three years (higher education: one year or less – Johnson et al., 2014a; K-12: two to three years – Johnson et al., 2014b). A systemic use of LA for improving teaching and learning in everyday practice, however, is still emerging (Siemens, 2014). For a their wide scale adaption, there is a need to carefully investigate the actual effects of integrating LA technologies in everyday educational practice (Chatti et al., 2012).

# 8. THE LEARNING ANALYTICS TOOLBOX: A COMPETENCE-BASED LEARNING ANALYTICS FRAMEWORK

In doing learning analytics a thorough understanding needs to be established on what is the objective, what needs to be known – and what data is most suitable to provide this information. Especially, since the amount and diversity of educational data is ever growing, new approaches are needed to exploit the informational potential residing in educational data (Long & Siemens, 2011). The LEA's BOX project aims at establishing a novel approach of competence-centred and theory-grounded LA that will help advancing LA research and application with respect to some of the challenges and problems outlined in sections 6 and 7 above.

The LEA's BOX approach will extend existing LA techniques by methods on the basis of two psycho-pedagogical frameworks, Competence-based Knowledge Space Theory (CbKST) and Formal Concept Analysis (FCA). CbKST originated in the field of adaptive learning environments and was elaborated to allow non-invasive formative skill assessment in line with LA ideas, as a basis for automated adaptation and support of learning experiences (Albert & Lukas, 1999; Heller et al., 2006). FCA originated from applied mathematics as an attempt to formalize concepts and concept hierarchies (Wille, 1982, 2005). In LEA's BOX a clearly learning-focused perspective is taken and a holistic framework for LA in terms of knowledge and competence-based modelling, structuring, assessment, and feedback is developed.

## 8.1 Basic Notions of CbKST

CbKST (Albert & Lukas, 1999; Heller et al., 2006) constitutes a framework for domain and learner knowledge and competence representation. CbKST is an extension of the purely behaviouristic Knowledge Space Theory (KST) (Doignon & Falmagne, 1985, 1999). Both, KST and CbKST constitute approaches for learner and skill modeling in learning environments and provide a sound theoretical basis for intelligent adaptation in technology-enhanced learning (Desmarais & Baker, 2012).

In KST a domain of knowledge is represented by a set of assessment problems; the subset of problems a person is capable of solving makes up the knowledge state of this individual. The knowledge domain is structured in terms of prerequisite relationships capturing dependencies among the problems. A prerequisite relation on a set of problems can be depicted by a so-called Hasse diagram, with ascending line segments indicating prerequisite relationships (see Figure 28(a) for an example domain of five problems). Consequently, not all subsets of problems will be potential knowledge states that are expected to be observable. The prerequisite relation establishes a quasi-order on the set of problems. The collection of knowledge states corresponding to the prerequisite relation builds the knowledge structure, which also contains the naïve knowledge state (covering no problem) and the expert knowledge state (covering the whole set of problems), as depicted in Figure 28(b). The knowledge states are naturally ordered by set inclusion and give rise to a number of meaningful learning paths (see Figure 28 (b)). Thus, given the knowledge state of a learner, a knowledge structure provides useful information, which learning content should be presented next, but also which previously learned material should be reviewed (Falmagne, Cosyn, Doignon, & Thiery, 2006). Furthermore, a knowledge structure can be exploited for realising an adaptive knowledge assessment, i.e. an assessment procedure that efficiently determines the current knowledge state of a learner by presenting him/her with only a subset of problems (Dowling & Hockemeyer, 2001; Falmagne et al., 2006). This can be realised by exploiting the structure inherent to the knowledge domain and selecting problems depending on previous answers of the learner. A knowledge structure can, thus, build the basis for creating personalised learning experiences, by tailoring the learning path to the learner's knowledge and aiming in closing the gap between current knowledge state and the desired knowledge state corresponding to the learning goal (Falmagne, 1993). In this way, Knowledge Space Theory constitutes a powerful basis for realising adaptive system behaviour in the context of technology-enhanced learning, allowing a tailoring of the learning process to the learner's knowledge, and has been successfully implemented in several e-learning environments (Albert, Hockemeyer, & Wesiak, 2002; Falmagne et al., 2006).

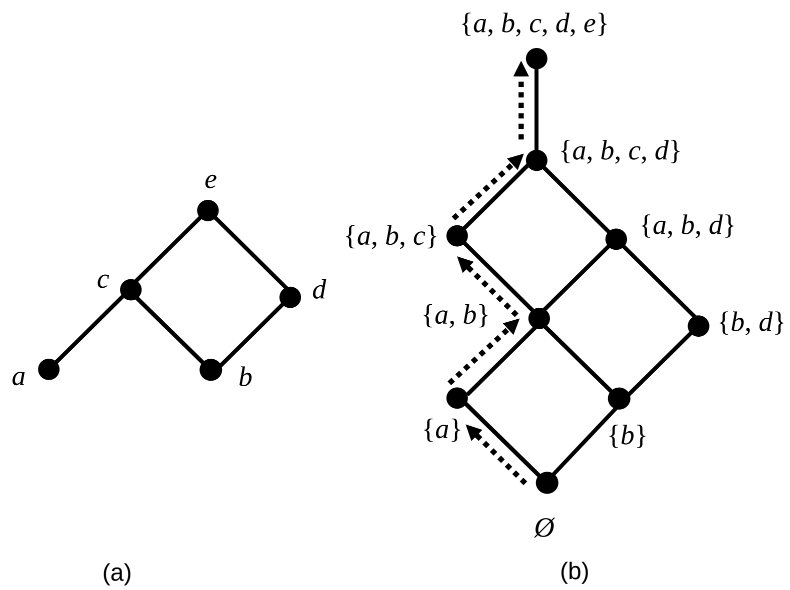


**Figure 28**: Example of a prerequisite relation (a) and the corresponding knowledge structure (b) with the dashed arrows representing a possible learning path.

Competence-based extensions of Knowledge Space Theory tried to incorporate competencies or skills into the theoretical framework and thus, to separate observable performance and underlying competence (Doignon, 1994; Düntsch & Gediga, 1995; Korossy, 1997; Albert & Held, 1994; Hockemeyer, 2003; Hockemeyer, Conlan, Wade, & Albert, 2003; Heller et al., 2006). These approaches aim at theoretically explaining the observed behaviour by considering underlying cognitive constructs in terms of skills. The basic idea of these approaches is to assume a basic set of skills or elementary competencies that describe abilities required for solving problems or taught by learning objects of a particular knowledge domain. The competence state of an individual is the collection of skills that he or she possesses. It is not directly observable but can be uncovered based on observable solution behaviour. Taking into account prerequisite relationships among skills, a competence structure can be established in analogy to a knowledge structure (Korossy, 1997, 1999). A competence structure constitutes the collection of all possible competence states that correspond to a prerequisite relation defined on the set of skills. By assigning to each problem the competence state(s) that is (are) sufficient for solving it (or, respectively, to each learning object the skills that it teaches), a knowledge structure on the set of problems (learning objects, respectively) is induced. This assignment of skills allows to identify a learner's available skills (i.e., his/her competence state) based on his/her observable performance and for deciding upon which learning objects are to be presented next, given a certain competence state, and thus builds the basis for generating personalised learning paths (Heller et al., 2006).

A common approach of defining skills in CbKST is by combining action/procedural and conceptual/declarative components (Marte, Steiner, Heller, & Albert, 2008). These skills can be related to existing educational taxonomies (e.g. Anderson & Krathwohl, 2001). The skill modelling approach of CbKST represents a competence-oriented and learner-centred approach to instruction and is in line with ideas of competence standardization for learning objectives and outcomes and competence-oriented instruction (e.g. BMUKK, 2012; European Communities, 2007, 2008; European Commission, 2012; Müller et al., 2013). CbKST has further nurtured research and development towards intelligent cognitive systems adapting to learners current competence state (e.g. Conlan, O'Keeffe, Hampson, & Heller, 2006; Albert, Hockemeyer, Kickmeier-Rust, Nussbaumer, & Steiner, 2012). The so-called microadaptivity approach (Augustin, Hockemeyer, Kickmeier-Rust, & Albert, 2011; Kickmeier-Rust & Albert, 2010) has been developed and applied in the context of game-based learning and integrates CbKST with theory of human problem solving (Newell & Simon, 1972) in order to model learners' behaviour and skills in problems solving during complex learning and assessment situations. This enables non-invasive assessment of learners' available and lacking skills by monitoring and interpreting their (inter)actions with a learning environment or serious game during problem solving, and the gathered assumptions on a learner's skills can be use for providing adaptive interventions and feedback tailored to the learner's needs, in line with ideas of formative assessment and LA (e.g. Kickmeier-Rust, Steiner, & Albert, 2011)

## 8.2 Basic Notions of FCA

Formal Concept Analysis (FCA), established by Wille (1982), aims to describe concepts and concept hierarchies in mathematical terms. The starting point of FCA is the definition of a so-called 'formal context' (learning domain). The formal context K is defined as a triple (G, M, I) with G as a set of objects which belong to the learning domain, M as a set of attributes which describe the learning domain´s, and finally, I as a binary relation between G and M. The relation I connects objects and attributes, i.e. $(g, m) \in I$ means the object g has the attribute m. The formal context K can be best read when depicted as a cross table, with the objects in the rows, the attributes in the columns and relations between them by ticking the according cells with an 'X' (see Table 12 for an example).

A formal concept is a pair (A, B) with A as a subset of objects and B as a subset of attributes. A is called the extension of the formal concept; it is the set of objects which belong to the formal concept. B is called the intension, it is the set of attributes which apply to all objects of the extension. The ordered and visualized set of all formal concepts is called the concept lattice $\mathcal{B}(K)$ (see Wille, 2005). This capability of visualizing the structure inherent among data is a strong feature of FCA. The concept lattice for the formal context example presented in Table 12 is depicted in Figure 29.

**Table 12:** Example formal context for the learning domain 'biotope' (based on Ganter and Wille, 1996).

| | ***Attributes*** | | | | |
|---|---|---|---|---|---|
| ***Objects*** | m1 *lives solely in the water* | m2 *is able to change location* | m3 *has limbs* | m4 *breastfeeds descendants* | m5 *applies photosynthesis* |
| Leech | X | X | | | |
| Frog | | X | X | | |
| Reed | X | | | | X |
| Goldfish | X | X | X | | |
| Elodea | X | | | | X |

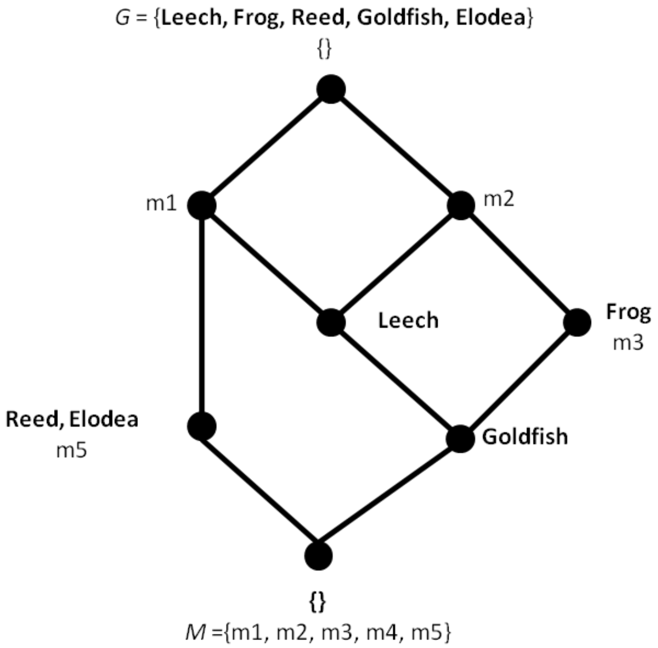


**Figure 29:** Concept lattice $\mathcal{B}$(K) of the learning domain 'biotope' as presented in **Table 12**.

Every node of the concept lattice represents a single formal concept. The extension of a particular formal concept is given by the objects, which can be reached by descending paths from that node. The intension is represented by all attributes, which can be reached by an ascending path from that node. For example, the node with the label 'leech' represents a formal concept with {leech, goldfish) as extension and {m1, m2} as intension. The concept lattice represents a conceptual clustering, with clusters represented simultaneously by objects and their attributes.

FCA has been applied in different areas, including knowledge management, text retrieval, data mining, economics, medicine, and software engineering (e.g. Ganter, 2009). The mathematisation of concepts and concept hierarchies through FCA provides a suitable approach of mathematically supporting humans in their logical thinking. For its success in terms of a multitude of application fields and the possibility to aggregate mathematical thought with other disciplines, FCA is also considered as having a function of 'transdsciplinary mathematics' (Wille, 2005). In this sense, Rush and Wille (1996) have shown that FCA can be connected to KST; more concretely, FCA can be used to represent the response patterns of individuals on a knowledge domain as a formal context. Through this linkage, methods of FCA become applicable to KST; for example it is possible to generate a knowledge space starting from a formal context through attribute exploration.

## 8.2 The LEA's BOX approach

A well-known dilemma in learning analytics is that of using top-down versus bottom-up analytic approaches. In LA commonly a bottom-up approach is taken, when it comes to data collection and analysis – learning data is gathered over a period of time and then analysed in order to extract valuable information patterns (Baker & Yacef, 2009). Instead of purely data-driven approaches of pattern recognition in learning-related data, there are increasingly claims for pursuing a more top-down like strategy, i.e. the idea that reasoning about data requires robust and well-elaborated psycho-pedagogical foundations. Chatti, Dyckhoff, and Schroeder, 2012, for example, argue that LA needs to start from a research question, in a first step, and the selection of suitable analysis methods should be done in a second step. Starting analytics from questions and psycho-pedagogical theory and models of teaching and learning, from conceptions of knowledge, of how learning takes place, and how learning success is manifested, is considered one of the main challenges in the emerging field of LA.

The competence-based LA framework of LEA's BOX consists in a hybrid approach building upon and harmonizing bottom-up and top-down procedures and aiming at realizing feasible, efficient, effective and pedagogically meaningful analysis and sense-making of learning related data. The general idea is to start with psycho-pedagogical considerations, information, and consultation to establish coarse, 'theory-driven' competence models in the sense of CbKST, which provide a meaningful representation of competences, with pedagogically and semantically rich underlying descriptions and the structure among them. These models serve as the cognitive basis and pedagogical hypothesis for carrying out LA. Through data-analytic methods in the tradition of FCA (for example as described in Ganter & Glodeanu, in press) models of abstract competences can be generated in a data driven way. Blending both approaches, the abstract, data-driven competence models may be enriched with pedagogical meaning of the theory-driven models, on the one hand. On the other hand, the coarse theory-driven models may be enriched by more fine-grained skills based on the data-driven results. In this way, combining, refining, and selecting most appropriate competence models shall be enabled. The established competence structures can be exploited for the purpose of continuous formative skill assessment, monitoring, reporting/feedback, and pedagogical intervention.

This hybrid, competence-based LA approach is currently being further elaborated and will be implemented in terms of a set of LA tools that will be part of the LEA's BOX platform. A service-based approach in LEA's BOX will provide flexibility in the (re-)use of the learning analytics tools developed in different educational environments and settings. In terms of data, for the analytical process in LEA's BOX a blend of various educationally relevant indicators from different learning systems and tools will be gathered and triangulated, to allow meaningful and nuanced analysis and interpretation. The information derived from LA will be reported back to teachers and learners. To this end, existing OLM approaches (Bull & Kay, 2010) will be enriched by visual and graph representations, such as competence structures depicted as Hasse diagrams or concept lattices known from CbKST and FCA (Kickmeier-Rust & Albert, 2013c), to come up with a set of dashboard visualizations that optimally support understanding of learner models and progress. These visualizations shall also serve the collection of human-contributed feedback and corrections to be used for learner model negotiation and potential further refinement of competence and LA models, corresponding to the claim of giving users the possibility to influence LA (Siemens, 2012) and the idea of continuous improvement of LA as part of the post-processing phase of the LA process (Chatti et al., 2012).

In LEA's BOX the issues of LA application in educational practice, educational relevance of analytics, and missing evidence of the impact of LA on teachers and learners, their behaviour, metacognition, teaching and learning experiences (e.g. Dyckhoff et al., 2013) are addressed. This is accomplished by a continuous engagement with educational practitioners and learners in partner schools Czech republic, Turkey, and Austria throughout the project, from requirements analysis to summative evaluations (cf. D5.1 Piloting and Evaluation Plan). In this way, a competence-based LA framework and LA toolbox shall be developed that are not only usable, but also useful for the target end users, enabling them to address relevant educational questions or objectives and empowering teaching and learning.

# 9. CONCLUSION & NEXT STEPS

The increasing use of a multitude of e-learning resources, educational software, learning tools, and the Internet in education produces a vast amount data related to learning. "This information provides a goldmine of educational data that can be explored and exploited to understand how students learn" (Romero & Ventura, 2012). The important question is how to exploit these different sets of educational data and make sense of them for assessment? Most of the learning management systems used today support basic level analytics, like average usage time, number of educational resources visited per learner etc. (Ferguson, 2013). For a holistic understanding of the learning process and progress, more sophisticated analyses are needed, and this is where approaches of LA come into play.

LA has enormous potential to transform the educational field, by providing new insights on how students learn and predicting learning. LA results may be used to better inform learners and teachers, and to implement better and smarter instructional methods and learning technologies. Analytics and data mining has developed relatively late in the educational field, compared to other fields, like business intelligence, biology, climate science. Considerable progress has been made in the last few years, for example in terms of a growing awareness that not all information relevant for learning can be captured in one data stream, but data of various types and sources needs to be integrated. In addition, LA research has started to model different constructs relevant to learning (like skill acquisition, affect, meta-cognition, collaboration, engagement), and analytics models continuously gain higher accuracy and quality through the continual improvement over time (Baker, 2014).

Nevertheless, there are still large steps to make to achieve broad adoption of LA technologies in educational practice. Apart from ensuring ICT infrastructure and skills, and openness and positive attitudes towards ICT and LA technologies, increasing educational and practical relevance and value, providing usable and useful LA tools, developing better standards for validation and impact assessment, and establishing appropriate ethical guidelines and procedures are the key objectives to be pursued in the next decade(s). Furthermore, LA technologies need to be incorporated in the workflow and day-to-day routine of educational practitioners and institutions. By further advancing LA in these directions, it may have an invaluable impact on the optimization and support of teaching and learning experiences and on the evolution and refinement of educational structures.

This document has provided a review the state of the art in the field of LA, given an overview of existing methods and tools, highlighted recent trends and current challenges and, in this way builds a solid starting point for the elaboration of the LA approach and technologies of LEA's BOX. Concretely, the outcome of this review feeds into the tasks engaged with the design and development of the general services and central executive (T3.2) and the design and development of the formal competence-based LA services (T3.3) and related upcoming WP3 deliverables. On the one hand, existing LA tools and approaches will serve to inspire and enrich the LEA's BOX platform. On the other hand, the applicability of CbKST and FCA approaches in LA will be advanced to serve as a cognitive basis for competence-centered analytics. The learning analytics framework currently under development constitutes an approach to effectively assess, monitor, report, and promote knowledge and competence, integrating educational data stemming from multiple activities and data sources. This approach is characterized by a combination of theory- and data-driven methods as a basis for competence-centred LA and will help to build a more comprehensive and accurate understanding of learning and progress that can be reported to students and instructors and will serve supporting learning, optimizing teaching, and refining LA methodologies itself. The conceptual research will be implemented in terms of a set of LA services and integrated into the global system.

# REFERENCES


Adomavicius, G. & Tuzhilin, A., “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, 734-749, 2005.

Albert, D. & Held, T., “Establishing knowledge spaces by systematical problem construction”. In D. Albert (ed.), Knowledge structures (78–112). New York: Springer, 1994.

Albert D. & Lukas J., “Knowledge spaces: Theories, empirical research, applications”. Mahwah: Lawrence Erlbaum Associates, 1999.

Albert, D., Hockemeyer, C., & Wesiak, G., “Current trends in eLearning based on knowledge space theory and cognitive psychology”. Psychologische Beiträge 44, 478-494, 2002.

Albert, D., Hockemeyer, C., Kickmeier-Rust, M. D., Nussbaumer, A., & Steiner, C. M., “E-learning based on metadata, ontologies and competence-based knowledge space theory”. In D. Lukose, A. R. Ahmad, & A. Suliman (Eds.), KTW2011. *Communications in Computer and Information Science* volume 295 of, pp. 24-36, Berlin. Springer, 2012.

Alcalá-Fdez, J., Sánchez, L., García, S., del Jesus, M.J., Ventura, S., Garrell, J.M., Otero, J., Romero, C., Bacardit, J., Rivas, V.M., Fernández, J.C., & Herrera, F., “KEEL: A software tool to assess evolutionary algorithms to data mining problems”. Soft Computing 13, 307-318, 2009.

Alcalá-Fdez, J., Fernandez, A., Luengo, J., Derrac, J., García, S., Sánchez, L., & Herrera, F., “KEEL Data-mining software tool: Data set repository, integration of algorithms and experimental analysis framework. Journal of Multiple-Valued Logic and Soft Computing 17, 255-287, 2011.

Alexa, M., & Zuell, C., “Text analysis software: Commonalities, differences and limitations: The results of a review”. Quality and Quantity, 34, 299-321, 200.

Ali, L., Asadi, M., Gasevic, D., Jovanovic, J., & Hatala, M., “Factors influencing beliefs for adoption of a learning analytics tools: An empirical study”. Computers & Education 62, 130-148, 2013.

Anderson, L.W. & Krathwohl, D.R (eds.), “A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives”. New York: Longman, 2001.

Arnold, K.E., “Signals: Applying academic analytics”. EDUCAUSE Review, 2010. Retrieved August 19, 2014 from http://www.educause.edu/ero/article/signals-applying-academic-analytics

Arnold, K.E. & Pistilly, M.D., “Course Signals at Purdue: Using learning analytics to increase student success.” In Proceedings of the International Conference on Learning Analytics and Knowledge (267-270). New York: ACM, 2012.

Arvaja, M. & Pöysä-Tarhonen, J., “Tracing discursive processes of shared knowledge construction in a technology-enhanced higher education setting”. Interactive Learning Environments, 2011.

Augustin, T., Hockemeyer, C., Kickmeier-Rust, M., & Albert, D., “Individualized skill assessment in digital learning games: Basic definitions and mathematical formalism”. IEEE Transactions on Learning Technologies, 4, 138-148, 2011.

Bahreini, K., Nadolski, R., & Westera, W., “Towards multimodal emotion recognition in e-learning environments”. Interactive Learning Environments, 1-16, 2014.

Baker, B., “A conceptual framework for making knowledge actionable through capital formation”. Dissertation. University of Maryland University College, 2007.

Baker, R.S., “Educational data mining: An advance for intelligent systems in education”. IEEE Intelligent Systems, 29, 78-82, 2014.

Baker, R., & Siemens, G., “Educational data mining and learning analytics”. In Sawyer, K. (Ed.) Cambridge Handbook of the Learning Sciences: 2nd Edition, in press.

Baker, R., Inventado, P.S. “Educational data mining and learning analytics”. In J.A. Larusson, B. White (Eds.) Learning Analytics: From Research to Practice (61-75). Berlin, Germany: Springer, 2014.

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

Baker, R.S.J.d., Habgood, M.P.J., Ainsworth, S.E., & Corbett, A.T., “Modeling the acquisition of fluent skill in educational action games”. In: Proceedings of User Modeling 2007 (17-26), 2007.

Bakharia, A. & Dawson, S., “SNAPP: A bird’s-eye view of temporal participant interaction”. In Proceedings of the International Conference on Learning Analytics and Knowledge 2011 (168-173). New York: ACM, 2011.

Barla, M., Bieliková, M., Ezzeddinne, A. B., Kramár, T., Šimko, M., & Vozár, O., “On the impact of adaptive test question selection for learning efficiency”. Computers & Education 55(2), 846–857, 2010.

Bastian M., Heymann S., & Jacomy M., “Gephi: an open source software for exploring and manipulating networks”. International AAAI Conference on Weblogs and Social Media. (2009).

Blikstein, P., “Multimodal learning analytics”. In Dan Suthers and Katrien Verbert (eds.), Proceedings of the Third International Conference on Learning Analytics and Knowledge (LAK '13) (102-106). New York: ACM, 2013.

BMUKK, "Kompetenzorientiertes Unterrichten an berufsbildenden Schulen. Grundlagenpapier” [Competence-oriented instruction at vocational schools. White paper]. Wien: Bundesministerium für Unterricht, Kunst und Kultur, 2012.

Brown, M., “Learning analytics: Moving from concept to practice”. EDUCAUSE Learning Initiative. July 2013. Retrieved July 7, 2014 from http://net.educause.edu/ir/library/pdf/ELIB1101.pdf

Brusilovsky, P., & Peylo, C., “Adaptive and intelligent Web-based educational systems”. International Journal of Artificial Intelligence in Education, 13, 159-172, 2003.

Buckingham Shum S., “Learning analytics. Dream, nightmare or fairydust?” Keynote task at Ascilite, Hobart, Tasmania, 2011. Retrieved July 28, 2014 from http://people.kmi.open.ac.uk/sbs/2011/12/learning-analytics-ascilite2011-keynote/

Buckingham Shum, S., “Learning analytics”. UNESCO Policy Brief. UNESCO Institute for Information Technologies in Education, 2012. Retrieved July 30, 2014 from http://iite.unesco.org/pics/publications/en/files/3214711.pdf

Buckingham Shum, S. & Ferguson, R., “Social learning analytics”. Educational Technology & Society 15, 3-26, 2012.

Bull, S. & Kay, J., “Open Learner Models”. In R. Nkambou, J. Bordeau and R. Miziguchi (eds.), Advances in Intelligent Tutoring Systems (318-338). Berlin: Springer, 2010.

Bull, S., Kickmeier-Rust, M., Vatrapu, R., Johnson, M.D., Hammermueller, K., Byrne, W. et al., “Learning, learning analytics, activity visualization and open learner model: Confusing?” In Hernández-Leo et al. (eds.), EC-TEL 2013. LNCS 8095 (532-535). Berlin: Springer, 2013.

Camilleri, V., de Freitas S, Montebello M, & McDonagh-Smith P., “A case study inside virtual worlds: Use of learning analytics for immersive spaces”. In: D. Suthers & K. Verbert (eds.), Proceedings of the 3rd International Conference on Learning Analytics and Knowledge (230–234). New York: ACM Press, 2013.

Campbell, J.P. & Oblinger, D.G., “Academic analytics”. EDUCAUSE Review, 42, 40-57, 2007.

Chang, K.-m., J. Beck, J. Mostow, & Corbett, A., “A Bayes Net Toolkit for student modeling in intelligent tutoring systems”. In Proceedings of the 8th International Conference on Intelligent Tutoring Systems, (104-113). Jhongli, Taiwan, 2006.

Chatti, M.A., Dyckhoff, A.L., Schroeder, U., & Thüs, H. “A reference model for learning analytics.” International Journal of Technology Enhanced Learning, 5, 318-331, 2012.

Chatti, M.A., Dyckhoff, A.L., Schroeder, U., & Thüs, H., “Forschungsfeld Learning Analytics. Learning Analytics Research Challenges.” i-com – Zeitschrift für interaktive und kooperative Medien, 1/2012, 22-25, 2012.

Conlan, O., O'Keeffe, I., Hampson, C., & Heller, J., “Using knowledge space theory to support learner modeling and personalization”. In T. Reeves & S. Yamashita (eds.), Proceedings of World

Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (1912-1919). Chesapeake: AACE, 2006.

Crow, M.C., “No more excuses. Michael. M. Crow on Analytics”. EDUCAUSE Review, 47, 14-22, 2012.

Davenport, T.H., Harris, J.G., & Morison, R., “Analytics at work. Smarter decisions. Better results”. Boston: Harward Business School Publishing, 2010.

Davis, F.D., Bagozzi, R.P., Warshaw, P.R., “User acceptance of com- puter technology: a comparison of two theoretical models”. Managment Science 35(8), 982–1003, 1989.

Desmarais, M.C. & Baker, R.S.J.d., “A review of recent advances in learner and skill modelling in intelligent learning environments”. User Model User-Adap Inter 22, 9-38, 2012.

Despotakis, D., Dimitrova, V., Lau, L., Thakker, D., Ascolese, A., & Pannese, L., “ViewS in user generated content for enriching learning environments: A semantic sensing approach”. In K. Yacef et al. (eds.), AIED 2013. LNAI vol. 7926 (121-130). Berlin: Springer, 2013.

De Liddo, A., Buckingham, Shum, S., Quinto, I., Bachler, M., & Cannavacciuolo, L., “Discourse-centric learning analytics”. In Proceedings of the International Conference on Learning Analytics and Knowledge, Banff, Alberta, 2011.

Denley, T., “DegreeCompass: A course recommendation system”. EDUCAUSE Review, September 2013. Retrieved October 21, 2014 from http://www.educause.edu/ero/article/degree-compass-course-recommendation-system

de Freitas, S., “Learning in immersive worlds. A review of game-based learning”. JISC E-learning programme, 2006. Retrieved March 1, 2013 from http://www.jisc.ac.uk/media/documents/programmes/elearninginnovation/gamingreport_v3.pdf

de Nooy, W., Mrvar, A., & Batagelj, V., “Exploratory social network analysis with Pajek. 2nd ed.” Cambridge University Press.

de Waard, I., “Adding learing analytics to your open online cloud course or MOOC”. Learning Solutions Magazine, October 2012. Retrieved October 22, 2014 from http://www.learningsolutionsmag.com/articles/1026/adding-learning-analytics-to-your-open-online-cloud-course-or-mooc

del Blanco, A., Serrano, Á., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B., “E-Learning standards and learning analytics. Can data collection be improved by using standard data models?” In Proceedings of the Global Engineering Education Conference (1255-1261). IEEE, 2013.

Doignon, J., “Knowledge spaces and skill assignments”. In G. H. Fischer & D. Laming (eds.), Contributions to mathematical psychology, psychometrics and methodology (111–121). New York: Springer, 1994.

Doignon, J.P. & Falmagne, J.-C., “Spaces for the assessment of knowledge”. International Journal of Man- Machine Studies 23, 175-196, 1985.

Doignon, J.P. & Falmagne, J.-C., “Knowledge spaces”. Berlin: Springer, 1999.

Dowling, C.E., & Hockemeyer, C., “Automata for the assessment of knowledge”. IEEE Transactions on Knowledge and Data Engineering 13, 451–461, 2001.

Drnek, M., Türker, A., Steiner, C., Hillemann, E., Kickmeier-Rust, M., “Deliverable D5.2: Focus groups and design report.” LEA’s BOX. Project no. 619762.

Drachsler, H. & Greller, W., “The pulse of learning analytics understandings and expectations from stakeholders”. In Proceedings of LAK12 (120-129). Vancouver: ACM, 2012.

Düntsch, I. & Gediga, G., “Skills and knowledge structures”. British Journal of Mathematical and Statistical Psychology 48, 9–27, 1995.

Dyckhoff, A.L., “Implications for learning analytics tools: A meta-analysis of applied research questions. International Journal of Computer Information Systems and Industrial Management Applications, 3, 594-601, 2011.

Dyckhoff, A.L., Lukarov, V., Muslim, A., Chatti, M.A, & Schroeder, U. “Supporting action research with learning analytics”. In: Proceedings of the International Conference on Learning Analytics and Knowledge (220–229). New York: ACM Press, 2013.

Duval, E., “Attention please! learning analytics for visualization and recommendation”. In Proceedings of the International Conference on Learning Analytics and Knowledge (9-17). New York: ACM, 2011.

Elias, T., “Learning analytics: Definitions , processes and potential”. Retrieved August 5, 2014 from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Essa, A. & Ayad, H., “Improving student success using predictive models and data visualisations”. In Research in learning technology. Supplement: ALT-C 2012 Conference proceedings (58-70). Association for Learning Technology, 2012. Retrieved August 19, 2014 from http://www.researchinlearningtechnology.net/index.php/rlt/article/viewFile/19191/pdf_1

EC, “Survey of Schools: ICT in Education. Benchmarking Access, Use and Attitudes to Technology in Europe’s Schools”. Final Report SMART-Nr 2010/0039, European Schoolnet & University of Liege, for the European Commission, DG Communications Networks, Content & Technology, 2013. ISBN 978-92-79-28121-1; doi:10.2759/94499

European Communities, “Key competencies for lifelong learning: European Reference Framework”. Luxembourg: Office for Official Publications of the European Communities, 2007.

European Communities, “Explaining the European Qualifications Framework for Lifelong Learning”. Luxembourg: Office for Official Publications of the European Communities, 2008.

European Commission, “Assessment of key competencies in initial education and training: Policy guidance”. Commission staff working document. SWD(2012) 371. Strasbourg, 2012.

European Union, “Smarter, greener, more inclusive? Indicators to support the Europe 2020 strategy. Indicators to support the Europe 2020 strategy. Luxembourg: Publicaitons Office of the European Union, 2013.

Falmagne, J., “Stochastic learning paths in a knowledge structure”. Journal of Mathematical Psychology 37, 489–512, 1993.

Falmagne, J.-C., Cosyn, E., Doignon, J.P., & Thiéry, N., “The assessment of knowledge, in theory and practice”. In R. Missaou & J. Schmidt (eds.), Formal Concept Analysis. 4th International Conference, ICFCA. Lecture Notes in Artificial Intelligence vol. 3874 (61-79). Berlin: Springer, 2006.

Ferguson, R., “Learning analytics: drivers, developments and challenges”. International Journal of Technology Enhanced Learning, 4, 304-317, 2012.

Ferguson, R., “Learning analytics for open and distance education”. In S. Mishra (Ed.), CEMCA EdTech Notes. New Delhi, India: Commonwealth Educational Media Centre for Asia (CEMCA), 2013.

Ferguson, R. & Buckingham Shum, S., “Learning analytics to identify exploratory dialogue within synchronous text chat”. In: Proceedings of the International Conference on Learning Analytics & Knowledge, Banff, Alberta, 2011.

Fernández-Gallego, B., Lama, M., Vidal, J.C., & Mucientes, M., “Learning analytics framework for educational virtual worlds”. Procedia Computer Science, 25, 443-447, 2013.

Ferrari, A., Brečko, B.N., Punie, Y., „DIGCOMP: A framework for developing and understanding digital competence in Europe”. E-learning papers 38, 1-15, 2014.

Fortenbacher, A., Klüsener, M., & Schwarzrock, S., “Ein generisches Datenmodell für Learning Analytics”. In C. Rensing & S. Trahasch, Proceedings of DeLFI Workshops 2014 (80-87). CEUR Workshop Proceedings vo. 1227, 2014.

Ganter, B., Stumme, G., & Wille, R. (ed.), “Formal concept analysis: Foundations and applications. LNAI vol. 3626. Berlin: Springer, 2009.

Ganter, B. & Glodeanu, C.V., “Factors and skills”. To appear in C.V. Glodeanu, M. Kaytoue, & C. Sacarea (Eds.), Formal concept analysis. LNAI vol. 8478, in press.

Giesbers, B., Rienties, B., Tempelaar, D., & Gijselaers, W., “Investigating the relations between motivation, tool use, participation, and performance in an e-learning course using web-videoconferencing”. Computers in Human Behavior 29(1), 285-292, 2013.

Govaerts, S., Verbert, K., Klerkx, J.,& Duval, E., "Visualizing activities for self-reflection and awareness", In Proceedings of the 9th International Conference on Web-based Learning (ICWL 2010). LNCS vol. 6483 (91-100). Springer, 2010.

Grann, J. & Bushway, D., “ Competency map: Visualizing student learning to promote student success”. In Proceeding of the International Conference on Learning Analytics and Knowledge (168-172). New York: ACM, 2014.

Greller, W. & Drachsler, H., “Translating learning into numbers: A generic framework for learning analytics”. Educational Technology & Society, 15, 42-57, 2012.

Groff, J., “Technology-rich innovative learning environments”. OECD Working Paper, 2013.

Harlen, W. & James, M., “Assessment and learning: Differences and relationships between formative and summative assessment”. Assessment in Education, 4, 365-379, 1997.

Heller, J., Steiner, C., Hockemeyer, C., & Albert, D., “Competence-based knowledge structures for personalised learning”. International Journal on E-Learning, 5, 75-88, 2006.

Hockemeyer, C., “Competence based adaptive e-learning in dynamic domains”. In F.W. Hesse & Y. Tamura (Eds.), The Joint Workshop of Cognition and Learning through Media–Communication for Advanced E-Learning (JWCL) (79–82). Berlin, 2003.

Hockemeyer, C., Conlan, O., Wade, V., & Albert, D., “Applying competence prerequisite structures for eLearning and skill management”. Journal of Universal Computer Science 9, 1428-1436, 2003.

Homer, B.D., “Introductory Talk to the Learning Analytics and Educational Data Mining Workshop”, CREATE Lab, New York University, April 2013.

Huang, S., & Fang, N., “Predicting student academic performance in an engineering dynamics course: A comparison of four types of predictive mathematical models”. Computers & Education 61, 133–145, 2013.

John, G. (2014, July 15). Bis zu 154.000 vorzeitige Bildungsabbrecher. Der Standard. Retrieved from www.derstandard.at

Johnson, L., Adams, S., & Cummins, M., “The NMC Horizon Report: 2012 Higher Education Edition.” Austin, Texas: The New Media Consortium, 2012

Johnson, L., Adams Becker, S., Estrada, V., & Freeman, A., “NMC Horizon Report: 2014 Higher Education Edition.” Austin, Texas: The New Media Consortium, 2014a

Johnson, L., Adams Becker, S., Estrada, V., and Freeman, A., ”NMC Horizon Report: 2014 K-12 Edition”. Austin, Texas: The New Media Consortium, 2014b.

Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Ludgate, H., “NMC Horizon Report: 2013 Higher Education Edition”. Austin, Texas: The New Media Consortium, 2013

Jovanović, J., "Generating context-related feedback for teachers." International Journal of Technology Enhanced Learning, 1, 47 – 69, 2008.

Kay, D., Korn, N., & Oppenheim, C., “Legal, risk and ethical aspects of analytics in higher education”. JISC CETIS Analytics Series: Vol. 1 No. 6, 2012. Retrieved October 22, 2014 from http://publications.cetis.ac.uk/c/analytics

Kaytoue, M., Marcuola, F., Napoli, A., Szathmary, L., & Villerd, J., “The Coron System”. In Suppl. Proceedings of the 8th International Conference on Formal Concept Analysis (ICFCA'10) (55-58). Agadir, Morocco, 2010.

Kickmeier-Rust, M.D., & Albert, D., “Micro adaptivity: Protecting immersion in didactically adaptive digital educational games”. Journal of Computer Assisted Learning, 26, 95-105, 2010.

Kickmeier-Rust, M. & Albert, D. “Learning analytics to support the use of virtual worlds in the classroom”. In: A. Holzinger & G. Pasi (eds.), Human-computer interaction and knowledge discovery in complex, unstructured, big data. LNCS vol. 7949 (358-365). Berlin: Springer; 2013a.

Kickmeier-Rust, M.D. & Albert, D., "Supporting formative assessment and appraisal by smart, competence-based, probabilistic systems.". International Conference on the Future of Education (FoE), Florence, 2013b.

Kickmeier-Rust, M.D. & Albert, D., "Using hasse diagrams for competence-oriented learning analytics". In A. Holzinger & G. Pasi (eds.), HCI-KDD 2013. LNCS vol. 7947 (59-64). Berlin: Springer, 2013c.

Kickmeier-Rust, M.D., Steiner, C.,M. & Albert, D., "Apt to adapt: Micro- and macro-level adaptation in educational games". In T. Daradoumis, S. Caballé, A. Juan & F. Xhafa (eds.), Technology-enhanced systems and tools for collaborative learning scaffolding". Studies in Computational Intelligence vol. 350 (221-238). Berlin: Springer, 2011.

Koedinger, K.R., Baker, R.S.J.d., Cunningham, K., Skogsholm, A., Leber, B., & Stamper, J., "A data repository for the EDM community: The PSLC DataShop". In C. Romero, S. Ventura, M. Pechenizkiy, & R.S.J.d. Baker (eds.), Handbook of educational data mining (43-56). Boca Raton, FL: CRC Press, 2010.

Koedinger, K.R. & Corbett, A.T., "Cognitive Tutors: Technology bringing learning sciences to the classroom". The Cambridge handbook of the learning sciences (61-78). Cambridge University Press, 2006.

Korossy, K., "Extending the theory of knowledge spaces: a competence-performance approach". Zeitschrift für Psychologie 205, 53-82, 1997.

Korossy, K., "Modeling knowledge as competence and performance". In D. Albert & J. Lukas (Eds.), Knowledge spaces: Theories, empirical research, applications (103-132). Mahwah: Lawrence Erlbaum Associates, 1999.

KnowledgeAdvisors, "Metrics that Matter and your LMS". Chicago: KnowledgeAdvisors, 2013. Retrieved October 22, 2014 from http://knowledgeadvisors.com/wp-content/uploads/2013/07/Metrics-that-Matter-and-Your-LMS.pdf

Kraan, W. & Sherlock, D., "Analytics tools and infrastructure". JISC CETIS Analytics Series: Vol. 1 No. 11, 2013. Retrieved October 22, 2014 from http://publications.cetis.ac.uk/c/analytics

Lee Goss, C. & Andren, K.J., " Dropout prevention". The Guilford practical interventions in the schools seriers. New York: The Guilford Press, 2014.

Lin, F.-R., Hsieh, L.-S., & Chuang, F.-T., "Discovering genres of online discussion threads via text mining". Computers & Education 52(2), 481-495, 2009.

Long, P., & Siemens, G., "Penetrating the fog. Analytics in learning and education." EDUCAUSE Review, 46, 30-40, 2011.

MacNeill, S., Campbell, L.M., & Hawksey, M., "Analytics for education". Journal of interactive Media in Education, 2012. Retrieved July 30. 2014 from http://jime.open.ac.uk/jime/article/view/2014-07

Marte, B., Steiner, C.M., Heller, J., & Albert, D., "Activity- and taxonomy-based knowledge representation framework". International Journal of Knowledge and Learning 4, 189-202, 2008.

Mayring, P., "Qualitative content analysis". Forum: Qualitative Social Research, 1(2), 2000. Retrieved May 14, 2008, from http://www.qualitative-research.net/fqs-texte/2-00/2-00mayring-e.pdf

Mazza, R., "Introduction to information visualization". London: Springer, 2009.

Mercer, N., "The analysis of classroom talk: Methods and methodologies". British Journal of Educational Psychology, 80, 1-14, 2010.

Miller, W.L., Baker, R.S., & Rossi, L.M., "Unifying computer-based assessment across conceptual instruction, problem-solving, and digital games". Technology, Knowledge, and Learning, 19, 165-181, 2014.

Mohamad, S.K. & Tasir, Z., "Eduational data mining: A review". Procedia – Social and Behavioral Sciences, 97, 320-324, 2013.

Müller, K., Gartmeier, M., & Prenzel, M., "Kompetenzorientierter Unterricht im Kontext nationaler Bildungsstandards". Bildung und Erziehung, 66, 127-144, 2013.

Newell, A. & Simon, H., "Human problem solving". Englewood Cliffs, NJ: Prentice-Hall, 1972.

Owston, R., “Teachers can make a difference: Professional development as a policy option for improving student learning with ICT”. CEIRIS Keris international Expert meeting on ICT and Educational Performance, South Korea, 2006.

Papamitsiou, Z. & Economides, A., “Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence”. Educational Technology & Society 17, 49-64, 2014.

Pardo, A., “Designing learning analytics experiences”. In J.A. Larusson & B. White (eds.), Learning analytics: From research to practice (15-38). New York: Springer, 2014.

Pistili, M.D., Willis, J.E., & Campbell, J.P., “Analytics through an institutional lens: Definition, theory, design, and impact”. In J.A. Larusson & B. White (eds.), Learning analytics: From research to practice (79-101). New York: Springer Science + Business Media, 2014.

Quinn, J. Drop-out and completion in higher education in Europe among students from under-represented groups. An independen report authored for the European Commission. Plymouth: Network of Experts on Social aspects of Education and Training (NESET), 2013. Retrieved Auguts 8 2014 from http://www.nesetweb.eu/sites/default/files/HE%20Drop%20out%20AR%20Final.pdf(2013).

Rayón Jerez, A., “Data analytics process in learning and academic analytics projects. Day 1: Data selection and capture [Powerpoint slides]. Retrieved October 22, 2014 from http://de.slideshare.net/alrayon/data-analytics01-data-selection-and-capture

Romero, C. & Ventura, S., “ Eductional data mining: A review of the state of the art.” IEEE Transactions on Systems, Man, and Cybernetics – Part C: Applications and Reviews, 40, 601-618, 2010.

Romero, C. & Ventura, S., “Data mining in education”. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery 3, 12-27, 2013.

Rosé, C., Wang, Y-C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F., “Analysing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning”. International Journal of Computer-Supported Collaborative Learning, 3, 237-271, 2008.

Rusch, A. & Wille, R., “Knowledge spaces and formal concept analysis”. In H. H. Bock & W. Polasek (Eds.), Data analysis and information systems: Statistical and conceptual approaches (427-436). Berlin: Springer, 1996.

Sao Pedro, M.A., Baker, R.S.J.d., Gobert, J., Montalvo, O., & Nakama, A., “Leveraging machine-learned detectors of systematic inquiry behavior to estimate and predict transfer of inquiry skill”. User Modeling and User-Adapted Interaction, 23, 1-39, 2013.

Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B., “Tracing a little for big improvements: Application of learning analytics and videogames for student assessment”. Procedia Computer Science, 15, 203-209, 2012a.

Serrano-Laguna, A., Torrente, J., Moreno-Ger, P., & Fernández-Manjón, B., “Application of learning analytics in educational videogames”. Entertainment Computing, 2014.

Serrano, A., Marchiori, E.J., del Blanco, A., Torrente, J., and Fernandez-Manjon, B.A. “A framework to improve evaluation in educational games”. In Proceedings of the 2012 IEEE Global Engineering Education Conference (1-8). IEEE, 2012b.

Schneider, D.K., “EduTech Wiki.” http://edutechwiki.unige.ch/en/Main_Page

Shaffer, D.W. & Arastoopour, G., “How research into epistemic network analysis might inform DCLA”. Paper presented at the 1st International Conference on Discourse-Centered Learning Analytics, Leuven, Belgium, 2013.

Shaffer, D. W., Hatfield, D, Svarovsky, G., Nash, P, Nulty, A., Bagley, E., Frank, K., Rupp, A., & Mislevy, R., “Epistemic Network Analysis: A prototype for 21st Century assessment of learning”. The International Journal of Learning and Media. 1(2), 33-53, 2009.

Shoukry, L., Göbel, S., & Steinmetz, R., “Towards mobile multimodal learning analytics”. In Proceedings of the Learning Analytics for and in Serious Games Workshop at EC-TEL 2014 (16). Graz, 2014.

Shute, V.J., "Focus on formative feedback." Review of Educational Research, 78, 153-189, 2008.

Shute, V.J. & Kim, Y.J., "Formative and stealth assessment". In J. M. Spector, M. D. Merrill, J. Elen, & M. J. Bishop (Eds.), *Handbook of Research on Educational Communications and Technology (4th Edition).* (pp. 311-323). New York, NY: Lawrence Erlbaum Associates, Taylor & Francis Group, 2013.

Shute, V. J., Ventura, M., Bauer, M. I., Zapata-Rivera, D., "Melding the power of serious games and embedded assessment to monitor and foster learning: Flow and grow". In U. Ritterfeld, M. Cody, & P. Vorderer (Eds.), *Serious games: Mechanisms and effects* (295-321). Mahwah, NJ: Routledge, Taylor and Francis, 2009.

Siemens, G., " Learning analytics: Envisioning a research discipline and a domain of practice. In Proceedings of the International Conference on Learning Analytics and Knowledge (4-8). New York: ACM, 2012.

Siemens, G., "The Journal of Learning Analytics: Supporting and promoting learning analytics research". Journal of Learning Analytics 1, 3-5, 2014.

Slade, S. & Prinsloo, P., "Learning analytics: ethical issues and dilemmas". American Behavioral Scientist, 57, 1509-1528, 2013.

Steiner, C.M., Kickmeier-Rust, M.D., & Albert, D., "Learning analytics and educational data mining: An overview of recent techniques". In Proceedings of the Learning Analytics for and in Serious Games Workshop at EC-TEL 2014 (6-15). Graz, 2014.

Stemler, S., "An overview of content analysis". Practical Assessment, Research & Evaluation 7, 2001. Retrieved May 14, 2008, from http://PAREonline.net/getvn.asp?v=7&n=17

Stephenson, S., Baker, R., Corrigan, S., "Towards building an automated detector of engaged and disengaged behavior in game-based assessments". Poster presented at the Annual Conference on Games+Learning+Society, 2014.

Taras, M., "Assessment – Summative and formative – Some theoretical reflections". British Journal of Educational Studies, 53, 466-478, 2005.

Terzis, V., & Economides, A. A., "The acceptance and use of computer based assessment." Computers & Education 56(4), 1032–1044, 2011.

Trčka, N. & Pechenizkiy, M., "From local patterns to global models: Towards domain driven educational process mining". In Proceedings of Ninth International Conference on Intelligent Systems Design and Applications (ISDA'09) (1114-1119), 2009.

Van Eck, R., "Digital game-based learning. It's not just the digital natives who are restless". Educause Review, 16-30, 2006.

Van Harmelen, M. & Workman, D., "JISC CETIS Analytics Series: Vol.1 No.3, Analytics for learning and teaching". University of Bolton, 2012. Retrieved August 19, 2014 from http://publications.cetis.ac.uk/2012/516

Ventura, M., Shute, V. J., Small, M., "Assessing persistence in educational games". In R. Sottilare, A. Graesser, X. Hu, & B. Goldberg (Eds.), *Design recommendations for adaptive intelligent tutoring systems: Learner modeling, Volume 2. (*93-101). Orlando, FL: U.S. Army Research Laboratory, 2014.

Verbert, K, Duval, E., Klerkx, J., Govaerts, S., & Santos, J.L., "Learning analytics dashboard applications". American Behavioral Scientist, 57, 1500-1509, 2013.

Wasson, B., & Hansen, C., "*Teaching with ICT and assessment literacy"*. Routledge, in press.

Wastiau, P., Blamire, R., Kearney, C., Quittre, V., Van de Gaer, E., & Monseur, C., "The use of ICT in education: a survey of schools in Europe." European Journal of Education, 48, 11-27.

Wille, R., "Restructuring lattice theory: An approach based on hierarchies of concepts". In I. Rival (ed.), Ordered sets (445–470). Dordrecht: Reidel, 1982.

Wille, R., "Formal concept analysis as mathematical theory of concepts and concept hierarchies". In B. Ganter , G. Stumme and R. Wille (eds.), Formal Concept Analysis (1-34), Berlin: Springer, 2005.

Willis, J.E., Campbell, J.P., & Pistilli, M.D., "Ethics, big data, and analytics: A model for application." EDUCAUSE Review, May 2013. Retreived October 28, 2014 from http://www.educause.edu/ero/article/ethics-big-data-and-analytics-model-application

Willis, J.E., "Learning analytics and ethics: A framework beyond utilitarianism". EDUCAUSE Review, August 2014. Retrieved October 28, 2014 from http://www.educause.edu/ero/article/learning-analytics-and-ethics-framework-beyond-utilitarianism

Witten, I.H., Frank, E., & Hall, M.A., "Data mining. Practical machine learning tools and techniques. 3rd ed." Burlington: Elsevier.

Zimmerman, B.J., "Becoming a self-regulated learner: An overview". Theory Into Practice, 41, 64-70, 2002.