Translating Learning into Numbers: A Generic Framework for Learning Analytics

Wolfgang Greller* and Hendrik Drachsler
Open University of the Netherlands, Valkenburgerweg 177, 6419AT Heerlen, The Netherlands // wolfgang.greller@ou.nl // hendrik.drachsler@ou.nl
*Correspondence author

ABSTRACT

With the increase in available educational data, it is expected that Learning Analytics will become a powerful means to inform and support learners, teachers and their institutions in better understanding and predicting personal learning needs and performance. However, the processes and requirements behind the beneficial application of Learning and Knowledge Analytics as well as the consequences for learning and teaching are still far from being understood. In this paper, we explore the key dimensions of Learning Analytics (LA), the critical problem zones, and some potential dangers to the beneficial exploitation of educational data. We propose and discuss a generic design framework that can act as a useful guide for setting up Learning Analytics services in support of educational practice and learner guidance, in quality assurance, curriculum development, and in improving teacher effectiveness and efficiency.

Furthermore, the presented article intends to inform about soft barriers and limitations of Learning Analytics. We identify the required skills and competences that make meaningful use of Learning Analytics data possible to overcome gaps in interpretation literacy among educational stakeholders. We also discuss privacy and ethical issues and suggest ways in which these issues can be addressed through policy guidelines and best practice examples.

Keywords

Learning analytics, Framework, Educational data mining, Ethics, Domain design, Data for learning

Introduction

In the last few years, the amount of data that is published and made publicly available on the web has exploded. This includes governmental data, Web2.0 data from a plethora of social platforms (Twitter, Flickr, YouTube, etc.), and data produced by various sensors such as GPS location data from mobile devices. In the wake of this, data-driven companies like Google, Yahoo, Facebook, Amazon, etc. are growing exponentially by commercially exploiting such data for marketing or in the creation of new personalised services. The new "data economy" empowers companies to offer an increasing amount of data products at little or no cost to their users (e.g., Google Flu Trends, bit.ly customised URLs, Yahoo Pipes, Gapminder.com). This growth in data also renewed the interest in information retrieval technologies. Such technologies are used to analyse data and offer personalised data products customised to the needs and the context of individual users.

It is already evident that data in combination with information retrieval technologies are not only the basis for the emergent data economy, but also hold substantial promises for use in education (Retalis et al., 2006; Johnson et al., 2011). One example of this is the research on personalisation with information retrieval technologies which has been a focus in the educational field for some time now (Manouselis et al., 2010). The main driver is the vision of improved quality, effectiveness, and efficiency of the learning processes. It is expected that personalised learning has the potential to reduce delivery costs while at the same time creating more effective learning experiences, accelerating competence development, and increasing collaboration between learners.

Not so long ago, for universities and companies alike, gathering data on their users met with substantial limitations in terms of cost, time requirements, scope, and authenticity of the data, as this was typically done using questionnaires or interviews with a selected representative number of stakeholders. The new data economy has made data collection very much an affordable activity. It is based on the highly economic electronic data mining of people's digital footprints and the automated analysis of behaviours of the entire constituency rather than sampling. Because data mining is not a separate act to normal user behaviour, the information retrieved is also highly authentic in terms of reflecting real and uninterrupted user behaviour. As such, data mining is more comparable to observational data gathering than to intrusive collection via direct methods. This will not make questionnaires and structured interviews obsolete, but it will greatly enhance our understanding and highlight possible inconsistencies between user behaviour and user perception (Savage and Burrows, 2007).

The proliferation of interactive learning environments, learning management systems (LMS), intelligent tutoring systems, e-portfolio systems, and personal learning environments (PLE) in all sectors of education produces vast amounts of tracking data. But, although these e-learning environments store user data automatically, exploitation of the data for learning and teaching is still very limited. These educational datasets offer unused opportunities for the evaluation of learning theories, learner feedback and support, early warning systems, learning technology, and the development of future learning applications. This leads to the importance of Learning Analytics (LA) being increasingly recognised by governments, educators, funding agencies, research institutes, and software providers.

The renewed interest in data science and information retrieval technologies such as educational data mining, machine learning, collaborative filtering, or latent semantic analysis in Technology-Enhanced Learning (TEL) reveals itself through an increasing number of scientific conferences, workshops and projects combined under the new research term Learning Analytics. Examples are the *1st Learning Analytics conference in Banff, Canada, 2011; the 4th International Conference on Educational Data Mining 2011 in Eindhoven, Netherlands; the 1st dataTEL workshop on Educational Datasets for Technology-Enhanced Learning at the Alpine-Rendez-Vous conference La Clusaz, France 2011; the 2nd International Conference on Learning Analytics and Knowledge (LAK12), Vancouver 2012; the 1st Workshop on Learning Analytics and Linked Data (LALD 2012)*; and more. Thus, the increasing amount of dedicated research events and publications make a meta-analysis of the domain timely and needed in order to establish a solid scientific basis which facilitates the development of new learner-oriented services.

Critical dimensions of learning analytics

Despite the great enthusiasm that is currently surrounding LA, it also raises substantial questions for research. In addition to technically-focused research questions such as the compatibility of educational datasets, or the comparability and adequacy of algorithmic and technological approaches, there remain several “softer” issues and problem areas that influence the acceptance and the impact of Learning Analytics. Among these are questions of data ownership and openness, ethical use and dangers of abuse, and the demand for new key competences to interpret and act on LA results. We shall point at these issues in more detail below. This means that the implementation of LA in learning processes requires to be carefully crafted in order to be successful and beneficial.

This necessity motivated us to identify six critical dimensions (soft and hard) of LA, which need to be covered by the design to ensure an appropriate exploitation of LA in an educationally beneficial way. By soft issues we mean challenges that depend on assumptions being made about humans or the society in general, e.g., competences or ethics. They are opposed by the hard challenges of the fact-based world of data and algorithms (cf. also the similar soft-hard distinction in Dron, 2011). In its coverage of soft issues, our framework differs from other, more workflow oriented models for LA, like that by Siemens (2011), although in his presentation he does acknowledge these as of concern. Rather than being a process model such as those collected in Elias (2011), we aim at a description framework that can later be developed into a domain model or ontology.

The critical dimensions highlighted here have been deduced from discussions in the emerging research community using a general morphological analysis (GMA) approach (cf. Ritchey, 2011). In this early formation stage of the LA community, scientific exchanges such as open online courses (MOOC) in Learning and Knowledge Analytics (LAK11, LAK12), or the above-mentioned events and congregations soon began to revolve around a number of key questions, like: Who is the target group for LA? What are we trying to achieve? How do we deal with privacy and data protection? These questions are naturally extended by other on-going debates such as the openness of data, which has been a topic for some time in the EDM and Open Linked Data domain, as well as technical and theoretical questions on achieving meaningful extraction of information from data.

Our chosen approach leading to the proposed framework consisted of a number of gathering and analysis processes. First, as a matter of opinion mining, we scanned the scientific interactions from proceedings and presentations of the conferences and working groups mentioned above. We conducted a brief literature review of abstracts in the field of Learning Analytics and Educational Data Mining. Additionally, we scanned the live discussions on the LA Google Groups (http://groups.google.com/group/learninganalytics and http://groups.google.com/group/LAK11), as well as the LAK11 MOOC (presentation chats and social networking exchanges). Furthermore, we looked back at recent RTD projects that contained elements of analytics and the questions and lessons they produced, e.g., the Language Technologies for Lifelong Learning project (http://www.ltfll-project.org) contained an analytics approach related to learner positioning and conceptualisation. Following these reviews, we applied cognitive mapping (Ackermann, Eden, and Cropper, 2004) for synthesising and sense making. We analysed these discussions and clustered them into the proposed six fields of attention, which we then presented as the first draft of the framework to a community of commercial and academic experts for evaluation and feedback at the *SURF seminar on Learning Analytics* (Eindhoven, 30-31 August 2011). The number six in the framework is not chosen for any particular reason, and other divisions are of course possible. However, we find the dissection into these six dimensions a useful and easy to follow domain orientation.

With the framework, we take the presumption that responsible designers of analytic processes will not only implement what is technically possible to do and legally allowed (or at least not prohibited), but to consider holistically the outcomes for stakeholders and, even more importantly, the consequences for the data subjects, i.e., the people supplying the data (cf. the section on stakeholders below). The framework intends to be a guide as much as a descriptor of the problem zones. Hence we refer to it as a "design framework" that can and should be used to design LA services from an inclusive perspective. We will argue below that this will help the transferability of LA approaches between different contexts of application and research.

Proposed design framework for learning analytics

Our proposed model for the domain and application of LA in figure 1 below considers six critical dimensions. Each of the dimensions can be subdivided into several instantiations falling into that dimension. For example, the generic "stakeholder" dimension can have instantiations (values) like "learners" and "teachers." The list of instantiations in the diagram is not exhaustive and can be extended on a case-by-case basis. To stay with the above example, commercial service providers and even automated agents could also function as stakeholders in a LA process. It is useful to note that through connecting various (and also multiple) different instantiations of each dimension, concrete use cases can be constructed. We call the dimensions "critical" in the sense that each of the six fields of attention is required to have at least one instantiation present in a fully formulated LA design. We realise, though, that some dimensions are vaguer than others in this respect.

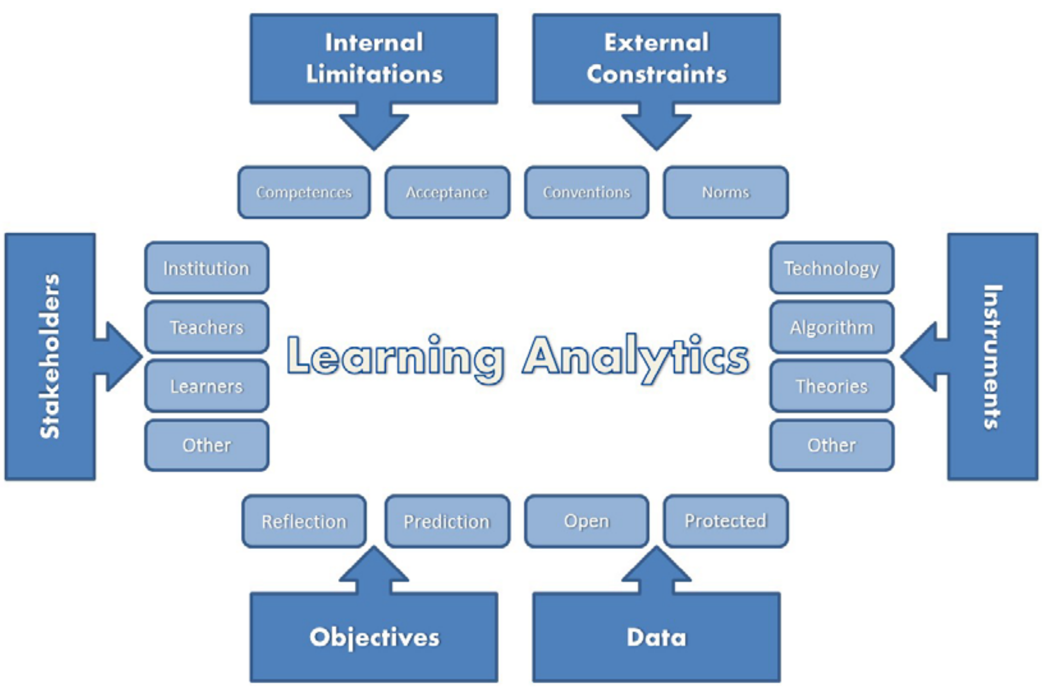

Figure 1. Critical dimensions of learning analytics

The six dimensions of the proposed LA framework are (cf. Figure 1): stakeholders, objectives, data, instruments, external constraints, and internal limitations. We will discuss each of these dimensions individually in the following and exemplify their instantiations and impact on the LA process and the benefits and opportunities they may determine. We will also elaborate apparent problem zones and limitations that may hinder any envisaged benefits.

Before embarking on the abstract dimensions in detail, we would like to illustrate the purpose and possible usage of the framework on the following sample use case, which is created out of a number of instantiations of the six dimensions. This specific example relates to conducting a social network analysis of students discussing in a forum using the SNAPP tool, based on the work by Dawson et al. (Dawson, 2008; Macfadyen & Dawson, 2010).

Table 1. Sample use case and values for dimensions

Dimension	Values
Stakeholders	*Data subjects*: a group of learners. *Data clients*: tutor, discussion moderator.
Objective	*Reflection*: Analyse student interactions in a forum discussion, identify network connections between students, and identify isolated students to bring them back into the discussion.
Data	*Protected dataset*: Student interactions and posts in the discussion forum of the LMS. *Relevant indicators*: Posts published, posts replied to. *Time scale*: what time frame is applied to the analysis?
Instruments	*Pedagogic theory*: socio-constructivist, hypothesis is that active participants in a discussion show better learning outcomes. *Technology*: Social Network Analysis (SNA), statistics. *Presentation*: network diagram visualisation, stats table.
External limitations	*Conventions: (1) Privacy*: is the analysis in accordance with privacy arrangements, are the students properly informed? *(2) Ethics*: What are the dangers of abuse/misguided use of the data? *Norms*: Are there e.g., legal data protection or IPR issues related to this kind of use of student data? *Time scale*: will the students still be able to benefit from the analytics outcome? Is the analysis post-hoc or just-in-time?
Internal limitations	*Required competences*: *(1) Interpretation:* Do the data clients have the necessary competences to interpret and act upon the results? Do they understand the visualisation or presentation of the information? *(2) Critical thinking:* Do they understand which data is represented and which data is absent? How will they use this information?

The above use case can be used (1) as a checklist when designing a purposeful LA process; (2) as a sharable description framework to compare context parameters with other similar approaches in other contexts, or for replication of the scientific environment. The framework allows an indefinite number of use cases with the respective value arguments.

Stakeholders

The stakeholder dimension includes *data clients* as well as *data subjects*. Data clients are the beneficiaries of the LA process who are entitled and meant to act upon the outcome (e.g., teachers). Conversely, the data subjects are the suppliers of data, normally through their browsing and interaction behaviour (e.g., learners). It is important to make this distinction in order to understand the impact of the process on individuals. In certain cases, the two types of stakeholder groups can be the same, as is the case if a LA application feeds back information to learners about their own learning rather than to inform the teacher, as would be a common case in informal learning scenarios. In the traditional learner-teacher scenario, the teacher would act as the data client, who receives information gathered from the data subjects, i.e., the learners.

As shown in the framework model (Figure 1), the main stakeholder groups of LA in formal learning situations are learners, teachers, and educational institutions. These may be expanded or substituted by other stakeholder groups, such as researchers, service providers, or governmental agencies. Each of the groups has different information needs and can be provided with tailored views on information using LA.

Information flow between stakeholders can best be exemplified with the common hierarchical model taken from formal education (Figure 2). What the diagram illustrates as an example is by which ways benefits might be obtained from LA. The pyramid encapsulates the academic layers of education and training institutions. In the most direct way, data analysis from the student level, e.g., via a LMS, can inform the above layer, in this case the teachers. Teachers can then use the analytics information to plan targeted interventions or adjust their pedagogic strategies. Institutions can, similarly, retrieve benefits from student and teacher data in order to provide staff development opportunities or to plan policies like quality assurance and efficiency measures. We also want to stress the major benefits LA offers for self-reflection on every level (cf. left side of the diagram). We would like to see institutions enabling and actively encouraging students to reflect on their learning data. But also teachers and institutions can gain new insights by reflecting on their performance. Not immediately involved in the learning processes, researchers (right of the diagram) could harvest data for the purpose of evaluating or innovating teaching processes or learning services. Finally (on top of the diagram), Government agencies may collect cross-institutional data to assess the requirements of Higher Education Institutes (HEI) and their constituencies.

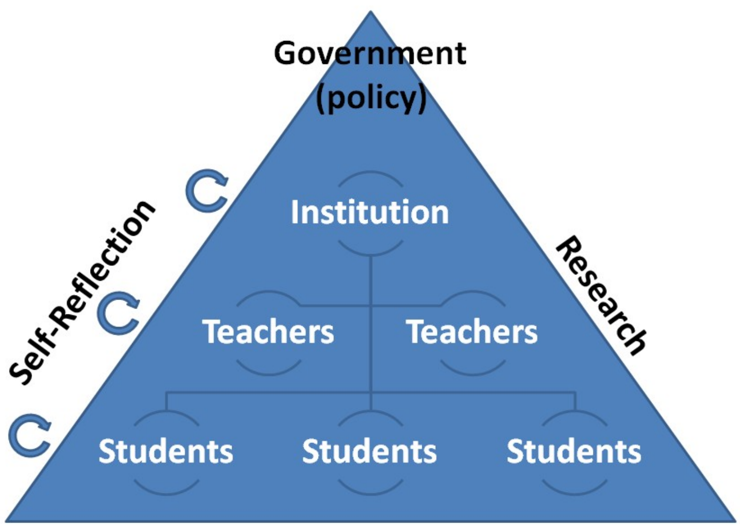

Figure 2. Information flow between LA stakeholders

Although they are the most widespread form in formal education, hierarchies are not the only flow models to describe where benefits can be retrieved. For example, peer evaluation using Personal Learning Environments (PLE) may be another information environment for LA. Peer environments also prevail in academic transactions like conferences or publications that are based on peer review systems. Practical examples for a horizontal peer-related information flow are the various scientific impact measures that exist, e.g., citation indexes. Equally, serious games can provide a non-hierarchical approach and/or team perspective to collaborative learning, e.g., how fast a team completed a level. In each of these, however, lie some issues of dependency and possible legal constraints (cf. further below).

Example opportunities for LA with respect to different stakeholder groups are:

> *Students* can be supported with specific learning process and reflection visualisations that compare their performance to the overall performance of a course. Furthermore, they can be provided with personalised recommendations for suitable learning resources, learning paths, or peer students (Gaviria et al., 2011).

> *Teachers* can be provided with course monitoring systems that inform them about knowledge gaps of particular pupils and thus enable them to focus their attention on those pupils. They can also harvest emergent group models that can lead to shared understanding of domain topics or processes for better curriculum design and on-the-fly adaptations.
>
> *Institutions* can monitor the performance of students regarding drop-out and graduation rate on a much finer granular level. In this way, they can evaluate their courses and improve outcomes of their courses.

Other stakeholders: We would like to emphasise that stakeholders need not be confined to formal education settings, but include all formal, non-formal, or informal environments, such as professional development (CPD). In these cases, the stakeholders are to be substituted by the relevant entities. For non-formal learning, for example, stakeholders would include a "learner" instantiation with (only) a self-reflection dimension in which feedback is mirrored back to the same person. In work-based learning, employees and line-managers may be the most common stakeholder groups involved. More notably, computer agents can also serve as stakeholders, for example as data clients that take further decisions on the learner's behalf or trigger an event (e.g., notification e-mail, recommendation of content or peer, etc.).

Objectives

The main opportunities for LA as a domain are to unveil and contextualise so far hidden information out of the educational data and prepare it for the different stakeholders (see above). Monitoring and comparing information flows and social interactions can offer new insights for learners as well as improve organisational effectiveness and efficiency. This new kind of information can support individual learning processes but also organisational knowledge management processes (Butler & Winne, 1995). We can distinguish two fundamentally different objectives: *reflection* and *prediction* (cf. Figure 1 above).

Reflection: Reflection is seen here as the critical self-evaluation of a data client as indicated by their own datasets in order to obtain self-knowledge. Wolf (2009) calls this process the "quantified self", i.e., self-observation and reacting to one's own performance log data. There already is a growing number of Personal Informatics Systems, i.e., human-computer interaction systems that support this process (Li & Forlizzi, 2010). However, reflection may also be seen as critical self-evaluation based on *other* stakeholders' datasets. This would especially be true if, for example, a teacher was led to reflect upon their teaching style as indicated by the datasets of their students. In the above hierarchical flow model (Figure 2), the higher order stakeholder would have the ability to utilise all the datasets from lower constituencies for their own reflection.

On an individual level, LA can support reflection of learning processes and offer personalised information on the progress of the learner (Govaerts et al., 2010). On the institutional level, LA can enhance monitoring processes and suggest interventions or activities for particular students. Greatest care should however be taken not to confuse objectives and stakeholders in the design of a LA process and not to let, e.g., economic and efficiency considerations on the institutional level dictate pedagogic strategies, as this would possibly lead to industrialisation rather than personalisation.

LA is a support technology for decision making processes. Therein also lies one of the greatest potential dangers. Using statistical analytic findings is a quantitative not a qualitative support agent to such decision making. We are aware that aligning and regulating performance and behaviour of individual teachers or learners against a statistical norm without investigating the reasons for their divergence may strongly stifle innovation, individuality, creativity and experimentation that are so important in driving learning and teaching developments and institutional growth.

Prediction: Apart from support for reflective practice, LA can also be used for predicting and modelling learner activities (Siemens, 2011; Verbert et al., 2011). This can lead to earlier intervention (e.g., to prevent drop-out), or to adaptive services and curricula. Using Machine Learning techniques, for example, learner profiles can be built dynamically and automatically, saving the learner filling in and maintaining profile data. In predictive outcomes lies currently much hope for efficiency gains in terms of establishing acts of automatic decision making for learning paths, thus saving teacher time for other more personal interventions. But prediction suffers potentially from big ethical problems (to which more further below), in that judgements about a person, whether originating from another human or a machine agent, if based on a limited set of parameters could potentially limit a learner's potential. For example, not every learner who has difficulties mastering subject level two, will automatically not master level three. We have to prevent re-confirming old-established prejudices of race, social class, gender, or other with statistical data, leading to restrictions being placed upon individual learners. Furthermore, there are limitations in the use of LA data as a means for supporting the learning process. Learning processes assume the leading role of the learner, rather than that of the teacher. However, the reliability of a LA-supported learner profile and its usefulness to the learners will remain questionable. For example, what LA data can be used in order to define whether a learning activity had a "high" or "low" impact on the learning process of learners, and at which points in the process itself? The diversity of learning makes it also problematic to judge which learning activity was of high value for learner A but of low value for learner B.

With respect to pedagogic theories, we would like to argue that LA does neither support nor ignore specific pedagogic theories, and as an abstract concept is pedagogically neutral. Indeed, we are of the opinion that LA can be used to evaluate different pedagogic strategies and their effects on learning and teaching through the analysis of learner data. This can be defined as a specific pedagogically oriented objective under the current dimension, but, as we will discuss further below, certain technologies are not pedagogically neutral and this will influence the analytics process in one way or another.

Educational data

LA takes advantage of available educational datasets from different Learning Management (LMS) and other systems. Institutions already possess a large amount of student data, and use these for different purposes, among which administering student progress and reporting to receive funding from the public authorities are the most commonly known. Linking such available datasets would facilitate the development of mash-up applications that can lead to more learner-oriented services and therefore improved personalisation.

LA strongly relies on data about learners and one of the major challenges LA researchers are facing is the availability of publicly available datasets to evaluate their LA methods. Most of the data produced in institutions is protected, and the protection of student data and created learning artefacts is a high priority for IT services departments. Nevertheless, similar to Open Access publishing and related movements, calls for more openness of educational datasets have already been brought forward (Drachsler et al., 2010). Anonymisation is one means of creating access to so-called Open Data. Recently, Verbert et al., (in press) presented a state of the art review of existing educational datasets. How open educational data should be, requires a wider debate (cf. section on legal constraints below), but, already in 2010, several data initiatives where started to make more educational data publicly available:

> *dataTEL challenge*—The first dataTEL challenge was launched as part of the first workshop on Recommender Systems for TEL (Manouselis et al., 2010), jointly organized by the 4th ACM Conference on Recommender Systems and the 5th European Conference on Technology Enhanced Learning (EC-TEL 2010) in September 2010. In this call, research groups were invited to submit existing datasets from TEL applications that can be used for LA research purposes and recommender systems for TEL.
>
> *dataTEL workshop*—The "Datasets for Technology Enhanced Learning" workshop was organised at the third STELLAR Alpine Rendez-Vous in March 2011. During this workshop, related initiatives that are collecting educational datasets, and apply these in data-driven learning applications were presented, and challenges related to privacy and data protection were discussed.
>
> *PSLC dataShop* (Stamper, 2011) offers an open data repository that provides access to a large number of educational datasets. *dataShop* has data from students derived from interactions with intelligent tutoring systems.
>
> *LinkedEducation.org* (Dietze et al., 2012) is another initiative that provides an open platform to promote the use of data for educational purposes. At the time of writing, five organizations have contributed datasets.

Despite these pioneering activities, it does, by comparison, still seem somewhat bizarre that in the commercial world with clicking the “register” button, the default access to all user data becomes owned by some company, whereas educational institutions operate on the default that everything is protected from virtually everyone.

Distinguishing educational data by access rights in open and protected datasets (Figure 1) is not as simple as it sounds. Because the technical systems producing and collecting data are typically owned by the institution, the easiest assumption would be that this data belongs to them. However, which employees of the institution exactly are included in the data contract between a learner (or their parents) and the educational establishment, is as yet unresolved. This poses severe constraints on inner-institutional research or wider institutional use. We will bring up some more legal consideration under the point on external constraints below.

Like in related research domains, LA datasets create a new set of challenges for research and practice. These include:

- A lack of common dataset formats like the suggested one from the CEN/ISSS PT social data group (cf. CAM Schema at: https://sites.google.com/site/camschema/home; and Wolpers et al., 2007).
- The need for version control and a common reference system to distinguish and point to different datasets.
- Methods to anonymise and pre-process data according to privacy and legal protection rights (Drachsler et al., 2010).
- A standardised documentation of datasets so that others can make proper use of it like that promoted by the data-seal-of-approval initiative (cf. http://www.datasealofapproval.org).
- Data policies (licences) that regulate how users can use and share certain datasets. For instance, the Creative Commons licensing rights could be considered as a standard way to grant permissions to datasets. DataCite (Brase, 2009) is an organization that enables to register research datasets and to assign licensing rights to them, so that the datasets can be referenced similar to academic articles.

From a technical point of view, idealised datasets probably remain the biggest challenge for analytics. This is to say that the assumption that datasets consist of context-free, meaningful and *only* meaningful data, is highly optimistic. In most natural settings, users “pollute” databases by producing erroneous or incomplete datasets. For example, teachers who want to see their students’ view on LMS courses often set themselves up as “test students” or create “test courses”. These are not always obvious, but need to be removed from the data to be analysed. Therefore empirical findings coming from a specific dataset are almost certainly affected by the context of data collection and processing.

Similarly, data collection often leads to “enmeshed identities” being used for analytics and prediction. A dataset cannot typically distinguish between a single individual and a shared presence in the learning space (group work on a single device). Students who often work together with others on shared devices (laptops, smartphone, lab space, etc.) produce enmeshed fingerprints in their educational data. This may lead to behaviours being attributed to a logged-in identity that may actually have originated from an “invisible” partner. Standardised documentation of datasets can be seen as paramount to raise awareness of this danger.

Additionally, from a pedagogic perspective, it remains an on-going challenge to formulate indicators from the available datasets that bear relevance for the evaluation of the learning process. The selection of specific data and their weighting (under the methods applied in the “instruments” dimension) against the real behaviour of students is of greatest importance, as is the process of relating behaviour pattern data to cognitive developments.

Instruments

Different technologies can be applied in the development of educational services and applications that support the objectives of educational stakeholders. LA takes advantage of so-called information retrieval technologies like educational data mining (EDM; cf. Romero et al., 2008), machine learning, or classical statistical analysis techniques (cf. Figure 1), but other techniques may also be considered relevant, e.g., social network analysis (cf. Buckingham & Ferguson, 2011) or Natural Language Processing (NLP).

Through these technologies, LA can contribute tailored information support systems to the stakeholders and report on demand. For instance, LA could be applied to develop a drop-out alert system. High drop-out rates are a challenging problem in education, especially distance education. Further research on LA can contribute to decrease the drop-out rate by developing e.g., a *Drop-out Analyser* that notifies the teacher of a course in time which students are in danger of falling behind or dropping out. This could be done by using LMS datasets and train a certain information retrieval technology (e.g., a Bayesian classifier) on the datasets to learn behavioural patterns of students that dropped out. Afterwards, the *Drop-out Analyser* could be applied on a follow-up online course and flag up students that show similar patterns. The teacher of the course could then intervene in an appropriate manner. Preliminary prototypes of such systems are already available, like the Blackboard Early Warning System.

Under the dimension "instruments" in our model (Figure 1), we also subsume conceptual instruments such as theoretical constructs, algorithms, or weightings, by which we mean different ways of approaching data. These ways in the broadest sense "translate" raw data into information. The quality of the output information and its usefulness to the stakeholders depend heavily on the methods chosen. Hildebrandt (2010), quite rightly, warns that "invisible biases, based on … assumptions … are inevitably embodied in the algorithms that generate the patterns".

Competing methods, technologies and algorithms applied to the same set of data, will result in different outcomes, and thus may lead to different consequences in terms of decision making based on these outcomes. LA designers and developers need to be aware that any algorithm or method they apply is reductive by nature in that it simplifies reality to a manageable set of variables (cf. Verbert et al., 2011).

External constraints

Many different kinds of constraints can limit the beneficial application of LA processes, some being "softer" than others. It has been suggested to us to identify them as ethical, legal, and social constraints, but also to feature organisational, managerial, and process constraints. This we find a useful subdivision of external limitations, but other divisions look equally logical. In the abstraction of the diagram above (cf. Figure 1), we propose the preliminary distinction of *conventions*, under which we count ethics, personal privacy, and similar socially motivated limitations, and, *norms* that are restricted by laws or specific mandated policies or standards. For reasons of space, we want to elaborate especially on the ethical aspects as this has grown into a field of much recent attention and debate (Bollier, 2010) and even spawned a collaborative effort in the Learning Analytics research community (Siemens, 2012).

New ethical and privacy issues arise when applying LA in education. These are challenging and highly sensitive topics when talking about educational datasets, as described in Drachsler et al. (2010). The feeling of endangered privacy may lead to resistance from data subjects toward new developments in LA. In order to use data in the context of LA in an acceptable and compliant way, policies and guidelines need to be developed that protect the data from abuse. Legal data protection may require that data is anonymised before it can be used. At the same time, as much openness of the datasets as possible is desirable (see paragraph on data above).

Personal data enjoys strong legal protection, differing by national laws and sometimes competing with other legal frameworks such as the Freedom of Information Act in the United Kingdom. We will not go into the legal details here, but did already above hint at the predicament that faces any data-related venture when using people's digital footprints. It concerns the lack of legal clarity with respect to data ownership. In current circumstances, data gathered about a person (before it is anonymised) belongs to the owner of the data collection tool, typically also the data client and beneficiary. Up till now, through direct intervention like questionnaires and sign-up processes, this was not a big problem to the data subjects. However, with the dramatic increase of ambient sensors and new technologies, such as location tracking or biometric face-recognition cameras, etc., more and more parts of individual behaviour are logged without the data subject's approval or even awareness. That being so, the ethic principle of "informed consent" (cf. AoIR, 2002) is very much under threat. The fundamental question legislators need to ask is: who does a person's life data belong to? (cf. also Hildebrandt, 2006). We believe that this question may in the near future become more and more elevated in importance and prominence.

On institutional level, educational and student data was traditionally handled separately, and is legally something of a blind spot. Registration data was kept and maintained by registry staff, IT data by IT staff, and learning data by academic staff. To use LA to its full potential, integration of available institutional datasets needs to happen. Universities, for instance, already collect and report socio-economic data such as students' post codes or ethnic and linguistic background. Institutions are even legally obliged by funding bodies to do so, but integrating this dataset with educational performance data, would be widely considered unethical or even illegal. As has been already mentioned above, the extent of a student's data contract with an institution and its individual staff representatives in different roles (teacher, administrator, secretary, researcher, IT support staff, Deans and management, etc.) needs to be urgently clarified. At the same time, privacy directives such as the *Data Seal of Approval* supported by the Dutch DANS institute (http://www.datasealofapproval.org) and related European data directives like the European Directive on data protection 95/46/EC (Directive, EU, 1995) need to be implemented.

Even where in compliance with the law, educational data can easily be abused for purposes inappropriate for educational institutions or for the data subjects (especially where minors are concerned). By principle, the more access to information about a data subject a data client has, the higher the responsibility is to use this information in a sensitive and ethical way. In an inspired article, Hildebrandt (2010) elaborates the ethically limited applicability of automated pattern recognition to the Law domain, but these limitations can be transferred just as easily to the domain of learning. Among the more obvious ethical risks are the exploitation of such data for commercial and similar purposes, or data surveillance issues (social sorting, cumulative disadvantages, digital stalking) and their ethical implications.

Ethics don't stop at the data gathering and integration. The realisation that we may encounter conflicts in values and interests in and through the analysis of people's behaviour needs to guide the post-analytic decision making process and the conclusions drawn from the approach. It is important to remind stakeholders of LA processes that data can be interpreted in many ways and lead to very different consequent actions. To give a drastic example, imagine being confronted with the insight that children from an immigrant background show reading difficulties, backed by supportive data analysis. This may lead to a wide ranging variety of responses, from developing extracurricular support mechanisms, to segregated classes, up to bluntly racist abuse of various kinds.

Ethics in LA may affect students and teachers alike, especially, where institutions aim to use LA to quality assure the performance of their teaching staff. Data can easily be abused as supporting evidence for exercising inappropriate pressures on data subjects to change otherwise perfectly acceptable or explainable performance behaviour. Institutions are therefore challenged with establishing a set of ethical policies and principles, together with, e.g., complaints procedures and safety nets that secure proper use of educational data in teaching and research. We find the ethical guidelines of the Association of Internet Researchers, AoIR (2002), a useful starting point in this respect, in that it has a purpose-oriented approach that supports ethical pluralism and respects the individual.

Another ethical consideration is the acceptance of divergence in the data constituency (AoIR, 2002). We already touched upon the danger that the result of algorithmic analysis, consequent policies and exercised pressures may aim at uniformity and at mainstreaming learning and teaching processes, thereby greatly harming creative processes and innovation that diverge from the statistical mean. It is one of the principal shortfalls of statistical prediction that it can only predict average behaviour not outliers. As such, LA provides no means of predicting exceptions to a rule, or exceptions to the exception rule.

Internal limitations

In complement to the environmental problem areas contained in the above "external constraints" section, we subsume a number of human factors that enable or may pose obstacles and barriers under the dimension heading "internal limitations." Prominent among these are *competences* and *acceptance*.

It is already becoming clear that the application of learning analytics requires new higher-order *competences* to enable fruitful exploitation in learning and teaching. In order to make LA an effective tool for educational practice, it is important to recognise that LA ends with the presentation of algorithmically attained results that require interpretation (Reffay & Chanier, 2003; Mazza & Milani, 2005). There are innumerable ways to present and to interpret data and base consecutive decisions and actions on it, but only some of them will lead to benefits and to improved learning. Basic numeric and other literacies, as well as ethical understanding are not enough to realise the benefits that LA has to offer. In a recent survey we conducted among LA experts, only 21% of the 111 respondents felt that learners would possess the required competences to interpret LA results themselves and determine appropriate actions/interventions from it (Drachsler & Greller, 2012). Therefore, the optimal exploitation of LA data requires some high level competences in this direction, but interpretative and critical evaluation skills (cf. Figure 1) are to-date not a standard competence for the stakeholders, whence it may remain unclear to them what to do as a consequence of a LA outcome or visualisation.

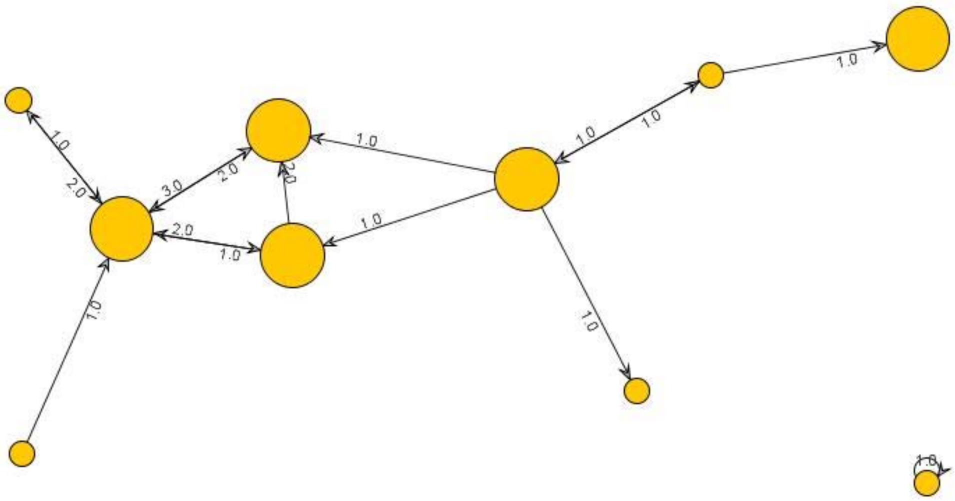

Figure 3. Social network diagrams may look attractive, but may not always be the best way to present information

Interpretation of LA results is often facilitated by enticing visualisations that are aimed to serve as a functional aid (Figure 3). One inherent danger that we perceive is that the simplicity and attractive display of data information may delude the data clients, e.g., teachers, away from the full pedagogic reality. This may negatively affect the pedagogic assessment and grading of a student's performance which should not be based alone on the visualisation of log files from a Learning Management System. To illustrate this danger take the example of marking student essays. An automated spell-check on orthographic mistakes presents itself as a quick and simple to interpret translation of the learner artefact into numbers. This makes it ideal for an efficient, cognitively effortless, and egalitarian grading mechanism. In an education environment that increasingly suffers from time constraints and calls for more efficiency in teacher activities, it is easily imaginable that the traditional qualitative assessment of essays gives way to such quick number crunching being over-proportionally reflected in student marks.

In our model (cf. Figure 1), we include among the key competences for LA, critical evaluation skills, because superficial digestion of data presentations can lead to wrong conclusions. It has to be strongly emphasised that data not included in the respective LA approach, is equally if not more important than the dataset that is included. To judge a learner's performance merely on, e.g., LMS quantitative data is like looking at a single puzzle piece. As learning is more and more happening in a lifelong and diverse ecosystem, an exclusive data view on single elements may provide a stimulus for reflection but not a sound basis for assessment.

The necessary competences notwithstanding, acceptance factors can further influence the application or decision making that follows an analytics process. This can, as is regularly seen in political debates, lead to blunt rejection of the results or applied methods from the constituency or parts thereof. In a learning context, ways to increase acceptance is vitally important also in order to produce usable outcomes. To get a better grasp on this issue, current scientific debate, therefore, should focus on empirical evaluation methods of learning analytics tools (Ali et al., 2012) and on advanced technology acceptance models (cf. Venkatesh & Bala 2008), inspired by the early work in this area (Davis, 1989, 1993). For LA, a revised technology acceptance model (TAM) could be an interesting approach to evaluate the emergent analytic technologies for all stakeholders described in our framework and also the needed implementation requirements to guarantee successful exploitation.

The place of pedagogy in the learning analytics framework

LA holds promises in the context of TEL by offering new methods and tools to diagnose learner needs and provide personalised instructions to better address these needs. It is not yet clear to what extent LA will lead to more personalised learning experiences rather than merely clustering people into behaviouristic "learner models" (e.g., as "outliers" of mainstream models). Consequently, more empirical evidence is needed to identify which pedagogic theory LA serves best. LA has been effectively used for behaviourist-instructivist style approaches (but see the critical reflection by Pardo & Kloos, 2011), but there is as yet little evidence for the support of constructivist approaches to learning (Duffy & Cunningham, 2001), where learning is seen as an active cognitive process in which learners construct their own concepts of the world around them. In LA, the latter is mostly inferred indirectly, by relating grades of learning outcomes with activities during the learning process (Dawson, 2012). In these correlations, it emerges that active students get better results. However, the role LA plays in this has not yet been conclusively demonstrated. Despite these questions, we would like to maintain that as knowledge and experiences vary considerably among learners, the diversity of learning can more effectively be addressed by LA methods than with current learning environments.

In our model, LA can work in support of a multitude of pedagogic strategies and learning activities as manifested and represented by the available data. This means we can only see pedagogies through the data. Because of this, we do not include them as part of the analytics process (Figure 1) but as *implicitly* contained in the input datasets that encapsulate the pedagogic behaviour of users. As we know, this behaviour depends a great deal on the platform and the pedagogic vision the developers built in (Dron & Anderson, 2011). For example, data from a content sharing platform will have a behaviourist/cognitivist pedagogy attached to the learner behaviour, since this is the pedagogic model underlying the technology. In any case, only the pedagogic patterns exhibited in the dataset can be analysed; and this will vary.

Additionally, pedagogy can be *explicitly* addressed in the goals and objectives that the LA designer sets ("objectives" dimension). The LA method ("instruments" dimension) will determine the outcome of the analysis and together with the interpretation applied may lead to a large variety of options for consequences and interventions. If such pedagogic interventions were applied, they would lead to new behaviours which, once again, could be analysed through the available data (Figure 4). In the graph below, we refer to pedagogic behaviour as learner/teacher behaviour that is motivated by didactic designs (learning designs). Pedagogic consequences, similarly, are adjustments to the didactic strategy or learning design based on the outcomes of the LA process.

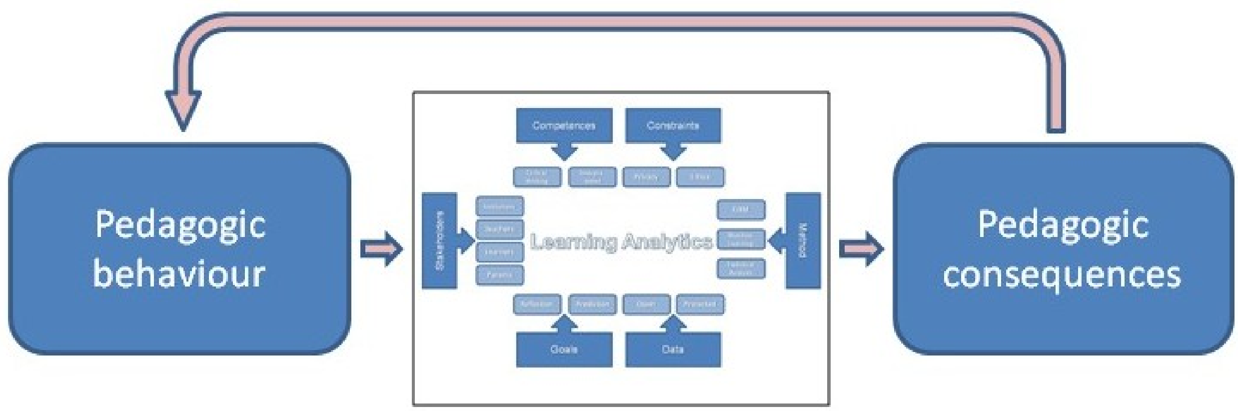

Figure 4. Learning analytics and pedagogy

A simple analogy would be boiling water in a pan. At any time (or continuously) you can immerse a thermometer and measure its temperature. The goal would be to determine whether you need to turn up the heat or not. The result of the analysis can then lead to the actions you want to take. The thermometer is only one method for such an analysis. An alternative would be to observe and wait until the water bubbles. Setting a benchmark (in the objectives design) can inform you when it is time for the teabag to go in. When applied to a learning process, immersing the thermometer into the water equates to the LA data gathering and analysis of learning in progress. It is here where learning is translated into numbers.

It should be noted that the pedagogic input factors are not confined to behaviour alone, but also include beliefs, (societal) values, and implicit theories of knowledge and learning. However, the LA application can only see these in the way they manifest themselves in the data. We also want to point at the possibility to apply computer agents to determine specific interventions. These could be as simple as sending a notification or recommendation to one of the stakeholders.

An example for pedagogic consequences is the following (Dawson, 2012): Using SNAPP as a tool to do a social network analysis (SNA) on discussion forums in a learning environment, the moderator or tutor might discover that certain changes in the moderation, the organisation, or the task, may lead to more or less engaged discussion among participants. In this way, the information gained through the LA process can support the fine-tuning of pedagogic effectiveness in a particular activity, depending on the desired learning outcome.

The model takes note that pedagogic success and performance is not the only thing that LA can measure. LA collects snapshots taken from educational datasets. These snapshots can be used to reflect or predict, in order to make adjustments and inform interventions, either by a human or by a system. Apart from offering efficiency benchmarking and business information for education providers, new support services for learning and more qualitative personal experiences can be achieved.

Conclusion and outlook

In summary, the proposed framework model in figure 1 above stresses the inherent connections between the six different dimensions and the impact of the analytics process on the end user and the data suppliers. If one of the parameters changes, the outcome and anticipated benefits will change. It is therefore our conviction that only the consideration of all six dimensions in the design process can lead to optimal exploitation of LA. Additionally, substantial work on new ethical guidance, data curation, and ownership needs to happen at universities and in legislation to reduce the risks connected to the application of LA and to protect the data subject, usually the learner.

Because of the inherent dependencies, we argue that all six dimensions are mandatory to be argumentatively present in a fully flexed LA design. We would, therefore, strongly welcome if application developers and researchers would not only make their technical environment known and open, but also describe the contextual environment and expectations from the users (e.g., required competences) along the lines of the framework. This would allow scientific comparison, replication, external analysis, and alternative contextualisation.

To validate the framework as both a descriptive approach as well as a guide to the design process of LA applications, we suggest evaluating the growing number of LA application showcases and testing for consistency in the descriptive values of the model. Additionally, we want to create a selected number of use cases that encompass the six dimensions and their instantiations.

LA is very much at the dawn of its existence and considered by many as one of the technological advances that will bring learning onto the next higher level. While we join in with this chorus of positive expectations, we are also aware that LA shows facets of a double nature: In its most optimistic outlook, learners will be provided with personal information about their current needs, while, at the same time, the educational system will be evolved from a "one-size-fits-all" approach into a highly personal competence-driven educational experience. But this view is not without flaws, because of the real dangers that the extended and organised collection of learner data may not so much bring added benefits to the individual, but instead provides a tool for HEIs, companies, or governments to increase manipulative control over students, employees, and citizens, thereby abusing LA as a means to reinforce segregation, peer pressure, and conformism rather than to help construct a needs-driven learning society. We therefore believe that it will be of critical importance for its acceptance that the development of LA takes a bottom-up approach focused on the interests of the learners as the main driving force.

LA has the potential for new insights into learning processes by making hitherto invisible patterns in the educational data visible to researchers and end users, and to enable development of new instruments for everyday educational practice. However, there are substantial uncertainties about the extent of impact LA will have on education and learning in general. The proposed framework model is motivated by the potential and opportunities that LA offers in its relevance for educational development and opportunities to personalise learning. While we agree with the Horizon report's forecast and its claim for a prosperous future of LA (Johnson et al., 2011), we also strongly feel that this development should not happen without a guiding framework that combines use of educational data with the protection of individuals and their learning.

Decisions based on LA are of concern, because they determine the usefulness and consequences for the stakeholders as well as the extent of its impact. Data analysis could have dramatic (and unwanted) consequences if not used with the necessary care. It is here where ethics play an enormously important role. Building of trust and confidence throughout the data constituencies has to be a priority from the start, and, here again, this proposed framework hopes to act as a useful guide.

One of the major questions in LA is the relation with theories of learning, teaching, cognition and knowledge. We hinted above at the opportunity that LA may support the evaluation of concrete didactic approaches which in turn may provide supportive evidence for particular pedagogic theories of learning and knowledge. At the same time, technologies are not pedagogically neutral; hence the evaluation will be influenced by the approach chosen. We consider this debate as an on-going one which will require further research and demonstration of applications and the impact they make on the process of learning.

It is still too early to base education fully on LA approaches alone, and we expect it never will be possible to do so. However, at the very least, opportunities this new discipline has to offer are to provide new support for learning activities and stimuli for reflection. In our opinion, it is these opportunities that LA should pursue.

References

Ackermann, F., Eden, C., & Cropper, S. (2004). Getting started with cognitive mapping. Retrieved from http://www.banxia.com/pdf/de/GettingStartedWithCogMapping.pdf

Ali, L., Hatala, M., Gasevic, D., & Jovanovic, J. (2012). A qualitative evaluation of evolution of a learning analytics tool. *Computers & Education, 58*(1), 470-489.

AoIR (Association of Internet Researchers) Ethics Working Committee, & Ess, C. (2002). Ethical decision-making and Internet research: Recommendations from the AoIR Ethics Working Committee. Retrieved from AoIR website: www.aoir.org/reports/ethics.pdf

Bollier, D. (2010). The promise and peril of big data. Washington, DC: The Aspen Institute.

Brase. J. (2009). DataCite – A global registration agency for research data. *Proceedings of the 2009 Fourth International Conference on Cooperation and Promotion of Information Resources in Science and Technology* (pp. 257–261). Washington, DC: IEEE Computer Society. doi: 10.1109/COINFO.2009.66

Buckingham Shum, S. and Ferguson, R. (2011). *Social learning analytics.* (Report No. KMI-11-01). Retrieve from Knowledge Media Institute, The Open University, website: http://kmi.open.ac.uk/publications/pdf/kmi-11-01.pdf

Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research, 65*, 245-281

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340

Davis, F. D. (1993). User acceptance of information technology: system characteristics, user perception and behavioral impacts. *International Journal of Man – Machine Studies, 38*, 475-487.

Dawson, S. (2012). Interpreting social networks: Informing teaching practice. Learning and Knowledge Analytics massive open online course (LAK12). Retrieved from https://sas.elluminate.com/p.jnlp?psid=2012-02-28.1237.M.0728C08DFE8BF0EB7323E19A1BC114.vcr&sid=2008104

Dawson S. (2008). A study of the relationship between student social networks and sense of community. *Educational Technology & Society, 11*(3), 224-238

Directive, EU (1995). 95/46/EC Protection of individuals with regard to the processing of personal data and on the free movement of such data. *Official Journal, L 281*, 0031 – 0050.

Dietze, S., Yu, H. Q., Giordano, D., Kaldoudi, E., Dovrolis, N., & Taibi, D. (2012, March). *Linked education: Interlinking educational resources and the web of data.* Paper presented at the 27th ACM Symposium on Applied Computing, Special Track on Semantic Web and Applications, Trento, Italy.

Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, …Wolpers, M. (2010). Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning. *Elsevier Procedia Computer Science, 1*(2), 2849–2858.

Drachsler, H. & Greller, W. (2012, April). *The pulse of learning analytics—Understandings and expectations from the stakeholders.* Paper presented at the Second International Conference in Learning Analytics (LAK12), Vancouver, Canada.

Dron, J. (2011). *Analytics: Soft and hard. Learning and knowledge analytics 2011 presentation* [Powerpoint slides]. Retrieved August 1, 2011, from http://www.slideshare.net/jondron/learning-analytics-soft-and-hard

Dron, J. & Anderson, T. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning, 80-97,* 80–97

Duffy, T. M. & Cunningham, D. J. (2001). Constructivism: Implications for the design and delivery of instruction. In D. H. Jonassen (Ed.), Handbook of Research for Educational Communications and Technology. New York, NY: Simon and Schuster.

Elias, T. (2011). Learning analytics: Definitions, processes and potentials. Retrieved from http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf

Gaviria, F., Glahn, C., Drachsler, H., Specht, M., & Gesa, R. F. (2011). Activity-based learner-models for learner monitoring and recommendations in Moodle. In C. D. Kloos et al. (Eds.), *Proceedings of the 6th European Conference on Technology-Enhanced Learning* (pp. 111-124). Heidelberg, Berlin: Springer-Verlag.

Govaerts, S., Verbert, K., Klerkx, J., Duval, E., (2010, December). *Visualizing activities for self-reflection and awareness.* Paper presented in the 9th International Conference on Web-based Learning, Shanghai University, China.

Hildebrandt, M. (2006). ’Privacy and Identity,’ privacy and the criminal Law. In E. Claes, A. Duff and S. Gutwirth (Eds.), *Antwerpen* (pp. 43–58). Oxford, UK: Intersentia. Retrieved from http://works.bepress.com/cgi/viewcontent.cgi?article=1005&context=mireille_hildebrandt

Hildebrandt, M. (2010, March). *The meaning and the mining of legal texts.* Paper presented at the Computational Turn in the Humanities, Swansea, Wales, UK. Retrieved July 11, 2011, from http://works.bepress.com/mireille_hildebrandt/27/

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

Li, I., Dey, A., & Forlizzi, J. (2010). A stage-based model of personal informatics systems. ACM Conference on Human Factors in Computing Systems, pp. 557–566. New York, NY: ACM. doi: 10.1145/1753326.1753409

Macfadyen, L., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), 588-599.

Manouselis, N., Drachsler, H., Vuorikari, R., Hummel, H., & Koper, R. (2010). Recommender systems in technology enhanced learning. In P. B. Kantor, F. Ricci, L. Rokach, & B. Shapira (Eds.), *Recommender Systems Handbook* (pp. 387-415). Berlin, Germany: Springer.

Mazza, R., & Milani, C. (2005, July). *Exploring usage analysis in learning systems: Gaining insights from visualisations.* Paper presented at the 12th International Conference on Artificial Intelligence in Education, Amsterdam, the Netherlands.

Pardo, A., & Kloos, C. D. (2011). Stepping out of the box: Towards analytics outside the learning management system. Proceedings of the 1st International Conference on Learning Analytics and Knowledge (pp. 163–167). New York, NY: ACM.

Reffay, C., & Chanier, T. (2003, June). *How social network analysis can help to measure cohesion in collaborative distance learning.* Paper presented at the International Conference on Computer Supported Collaborative Learning, Bergen, Norway.

Retalis, S., Papasalouros, A., Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). *Towards networked learning analytics—A concept and a tool.* Proceedings of the Networked Learning Conference 2006. Retrieved from http://www.networkedlearningconference.org.uk/past/nlc2006/abstracts/Retalis.htm

Ritchey, T. (2011). General Morphological Analysis—A general method for non-quantified modelling. In T. Ritchey (Ed.), *Wicked Problems—Social Messes.* Retrieved from http://www.swemorph.com/pdf/gma.pdf

Romero, C., Ventura, S. Espejo, P. G., & Hervs, C. (2008). Data mining algorithms to classify students. In R. de Baker, T. Barnes, J. Beck (Eds), *Proceedings of the 1st International Conference on Educational Data Mining* (pp. 8–17). Retrieved from http://www.educationaldatamining.org/EDM2008/uploads/proc/1_Romero_3.pdf

Savage, M. & Burrows, R. (2007). The coming crisis of empirical sociology. *The Journal of The British Sociological Association, 41*(5), 885–899. doi: 10.1177/0038038507080443

Siemens, G. (2011). Learning analytics: A foundation for informed change in higher education. Retrieved 1 August, 2011, from http://www.slideshare.net/gsiemens/learning-analytics-educause

Siemens, G. et al. (2012). Learning analytics: Guidelines for ethical use. Shared effort of the learning analytics research community. Retrieved March 23, 2012, from http://bit.ly/wMDmLW

Stamper, J. (2011). *EDM and the 4th paradigm of scientific discovery* [Powerpoint Slides]. Retrieved 1 August, 2011, from https://pslcdatashop.web.cmu.edu/about/edm_stamper_2011.html

Venkatesh, V., & Bala, H. (2008), Technology acceptance model 3 and a research agenda on interventions. *Decision sciences, 39*(2), 273–315.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R., & Duval, E. (2011, February). *Dataset-driven research for improving recommender systems for learning.* Paper presented in 1st International Conference Learning Analytics & Knowledge, Banff, Alberta, Canada.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (in press). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society*.

Wolf, G. (2009, July 17). Know thyself: Tracking every facet of life, from sleep to mood to pain, 24/7/365. *Wired Magazine*. Retrieved from http://www.wired.com/medtech/health/magazine/17-07/lbnp_knowthyself?currentPage=all

Wolpers, M., Najjar, J., Verbert, K., & Duval, E. (2007). Tracking actual usage: The attention metadata approach. *Educational Technology & Society*, *10*(3), 106-121.