Open Research Online

The Open University's repository of research publications
and other research outputs

Learning analytics: drivers, developments and challenges

Journal Item

oro.open.ac.uk

Learning analytics: drivers, developments and challenges

Rebecca Ferguson

Institute of Educational Technology,
The Open University, Walton Hall,
Milton Keynes, UK
Email: rebecca.ferguson@open.ac.uk

Abstract: Learning analytics is a significant area of technology-enhanced learning that has emerged during the last decade. This review of the field begins with an examination of the technological, educational and political factors that have driven the development of analytics in educational settings. It goes on to chart the emergence of learning analytics, including their origins in the 20th century, the development of data-driven analytics, the rise of learning-focused perspectives and the influence of national economic concerns. It next focuses on the relationships between learning analytics, educational data mining and academic analytics. Finally, it examines developing areas of learning analytics research, and identifies a series of future challenges.

Biographical notes: Rebecca Ferguson is a full-time Research Fellow in the UK Open University's Institute of Educational Technology, where her work is focused on how people learn together online. She works as a research lead on the *SocialLearn* team, developing and researching initiatives to improve pedagogical understanding of learning in online settings and to design analytics to support the assessment of learning in these settings.

1 Introduction

Learning analytics is a fast-growing area of Technology-Enhanced Learning (TEL) research. It has strong roots in a variety of fields, particularly business intelligence, web analytics, educational data mining and recommender systems. Its strong connections to these fields have meant that researchers and practitioners have approached it from a range of perspectives and must now work together to identify not only the goals that can be achieved using learning analytics but also what must be done in order to attain these goals. This paper and its companion piece, which sets out a reference model for learning analytics (Chatti et al., this issue) are original because they provide reviews of the development of learning analytics, which have been lacking until now, they examine the context within which further work will take place and they identify the challenges that lie ahead.

This paper has three aims:

1   to identify the drivers behind learning analytics, and its key points of reference;

2   to account for the emergence of the field of learning analytics over the past decade and for its divergence from other fields, particularly academic analytics and Educational Data Mining (EDM);

3   to make a significant contribution to future research by using these findings to identify the challenges that learning analytics must now address.

The development of the field is presented here in a broadly chronological structure, demonstrating the increasingly rapid pattern of development as new drivers emerge, new fields are appropriated and new tools developed. Tracing the development of learning analytics over time highlights a gradual shift away from a technological focus towards an educational focus; the introduction of tools, initiatives and methods that are significant in the field today; and the issues that have not yet been addressed.

This paper employs the definition of learning analytics set out in the call for papers of the first international Conference on Learning Analytics and Knowledge (LAK 2011) and adopted by the Society for Learning Analytics Research (SoLAR):

> Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs.

As it stands, this definition could be taken to cover the majority of educational research, but it is typically coupled with two assumptions: that learning analytics make use of pre-existing, machine-readable data, and that its techniques can be used to handle 'big data', large sets of data that would not be practicable to deal with manually.

As this is a newly defined field, papers relating to learning analytics draw on a diverse range of literature from fields including education, technology and the social sciences. In order to identify key points of reference, this paper focuses on the most cited papers and authors in this area. These were derived from submissions to the second international Conference on Learning Analytics and Knowledge (LAK 2012). This conference, organised by SoLAR, was the largest gathering of learning analytics specialists to date; it brought together 210 participants in Vancouver and many others submitted or co-authored papers. It therefore provided the most comprehensive overview to date of work in the field. As a member of the programme committee, the author had access to over 70 papers submitted to the conference. The bibliographies from these papers were collated, which produced a list of 1337 references. Publications that received four or more references (these 12 papers appear in bold in the bibliography) and all authors who were referenced five or more times (the names of these 20 authors appear in bold in the bibliography, and in bold italics when they authored one of the 12 papers with four or more references) were taken to be key reference points for the field, and it is those works and authors that form the backbone of this review, with references to other material employed to explain and contextualise their contribution.

There are limitations to this approach. These papers were all written in English, and their authors had access to sufficient money to attend conferences. This study therefore under-represents the learning analytics work carried out by non-English speakers and those from poorer countries. The focus on learning analytics sets aside related work in educational data mining, but this is covered elsewhere in this issue (see Chatti et al.). The

method took into account all references and thus includes 'grey literature' – including technical reports, Horizon Reports and some Educause material – that has not been subjected to peer review. Where possible, these have been supplemented by subsequent, peer-reviewed versions of the same material. The choice of dataset means that even the most significant papers presented at the LAK12 conference are not cited here, as none have had time to establish themselves as key reference points.

Section 2 addresses the first aim of this paper by identifying the technological, pedagogical and political/economic drivers that have motivated the emergence and development of learning analytics and related fields. The following sections address the paper's second aim and account for the emergence of the field of learning analytics in terms of these drivers, using a broadly chronological framework to chart the shift from data-driven analytics (Sections 3 and 4) towards learner-focused analytics (Section 5), influenced by political and economic concerns (Section 6). As education's interest in big data expands (Section 7), learning analytics is established as its own field (Section 8). The final section draws on those that precede it in order to meet the paper's third aim, identifying challenges that must be addressed in future.

2 Factors driving the development of learning analytics

2.1 Big data

Society is faced with the growing challenge posed by 'big data', 'datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyse' (Manyika et al., 2011, p.1). Businesses employ analytics to extract value from such datasets, using them to drive recommendation engines, identify patterns of behaviour and develop advertising campaigns. The widespread introduction of virtual learning environments (VLEs) – also known as learning management systems (LMSs) – such as *Blackboard* and *Moodle* has meant that educational institutions deal with increasingly large sets of data. Each day, their systems amass ever-increasing amounts of interaction data, personal data, systems information and academic information (Mazza and Milani, 2004; Romero et al., 2008). Although student-tracking capabilities are typically included as generic software features, the depth of extraction and aggregation, reporting and visualisation functionality of these built-in analytics has often been basic or non-existent (Dawson, 2009). In addition, significant amounts of learner activity take place externally, and so records are distributed across a variety of different sites with different standards, owners and levels of access. The first driver, then, is a technical challenge: *How can we extract value from these big sets of learning-related data?*

2.2 Online learning

The rise of big data in education mirrors the increase in take-up of online learning. Learning online offers many benefits, but it is also associated with problems. Students may feel isolated due to lack of contact with teachers or peers; they may become disorientated in the online space, experience technical problems or lose their motivation (Mazza and Dimitrova, 2004). At the same time, teachers lack the visual cues that can signal when students are not sufficiently challenged, when they are bored, confused, overwhelmed or simply absent. They may also struggle to interpret and evaluate the

learning and quality of participation of individuals when this is buried within hundreds of student contributions to discussions that have lasted several weeks (Dringus and Ellis, 2005). The second driver is therefore an educational challenge: *How can we optimise opportunities for online learning?*

2.3 Political concerns

There is increasing demand for educational institutions to measure, demonstrate and improve performance. This demand is evident in many countries (Campbell et al., 2007; EU Expert Group, 2010). In the context of analytics, it has been most clearly articulated within the USA, where the government aims to increase the overall educational attainment of the population and has been prepared to invest billions of dollars in order to achieve this (Norris et al., 2008). The third driver is therefore a political/economic challenge: *How can we optimise learning and educational results at national or international levels?*

2.4 Who benefits?

These three drivers draw attention to three different interest groups: governments, educational institutions and teachers/learners. Although the interests of all three groups overlap, they require analytics work on different scales and at different granularities. The choice of target audience therefore affects how researchers conceptualise problems, capture data, report findings, act on their findings and refine their models. As the following sections show, the field of analytics changes and develops as the balance between these three drivers and three interest groups shifts.

3 Origins in the twentieth century

Before the widespread emergence of online learning or big data, educational institutions were already involved in institutional research and evaluation. In 1979, the Survey Research Department of The Open University in the UK could reflect on ten years monitoring the progress of their many thousands of distance students, course by course, at several stages in the academic year (McIntosh, 1979). Even at this early date, McIntosh wrote of a 'data explosion', with the richness and profusion of available data acting as a barrier to its use.

Research was not confined to individual institutions; when Tinto published his studies of factors affecting student persistence, he was able to draw on a wide-ranging database of studies gathered over 20 years and covering a variety of institutional settings and types of students (Tinto, 1997). His synthesis of work on persistence, and his emphasis on the importance of academic and social integration, proved influential as institutions looked to analytics to address the problem of rates of student drop out.

At the time, online learning and interaction were in their infancy, with only limited institutional take-up of communication systems such as *FirstClass* and VLEs such as *TopClass* and *WebCT*. Early pioneers had found that communities could exist on the net (Rheingold, 1993), but it was only slowly becoming clear that collaborative learning could take place online (Dillenbourg, 1999). Understanding of how learning took place online was not yet sufficiently advanced to prompt the development of pedagogically inspired, learner-driven analytics.

4 Early 21st century: data-driven analytics

During the next three years, the situation changed. The emergence of the second-generation web, the 'read/write web', opened up new possibilities for collecting web content from diverse sources, processing it and exchanging the results with other programs (Berners-Lee et al., 2001). There was also rapid take-up of VLEs – UK figures suggest that 7% of higher education institutions were using a VLE in 1994, 40% in 2001 and over 85% by 2003 (Britain and Liber, 2004).

With extensive datasets increasingly available for analysis, the field of educational data mining gradually emerged (Romero and Ventura trace its origins to 1995, but only cite two papers pre-2000). Overall, data mining is a field of computing that applies a variety of techniques (for example, decision tree construction, rule induction, artificial neural networks, instance-based learning, Bayesian learning, logic programming and statistical algorithms) to databases in order to discover and display previously unknown, and potentially useful, data patterns (Chatti et al., this issue; Romero and Ventura, 2007). Educational data mining is a sub-set of this field and is

> concerned with developing methods for exploring the unique types of data that come from educational settings, using those methods to better understand students, and the settings which they learn in (www.educational datamining.org).

EDM emerged from the analysis of logs of student-computer interaction and, until 2005, relationship-mining methods were the most prominent type of EDM research, followed by prediction methods (Baker and Yacef, 2009). Despite the data-driven basis of the field, it has always had a strong emphasis on learning and teaching. Zaïane's 2001 paper, the most-cited in EDM (Romero and Ventura, 2007), identified the goal of educational data mining as 'turning learners into effective better learners' with research focused on

> data mining and machine learning techniques that could be used to enhance web-based learning environments for the educator to better evaluate the learning process, as well as for the learners to help them in their learning endeavour (Zaïane, 2001).

This perspective stands in contrast to early use of the term 'learning analytics' to refer to business intelligence about e-learning (Mitchell and Costello, 2000).

5 The rise of learning-focused perspectives

Alongside the data-driven approach to analytics, socially and pedagogically driven approaches to analytics began to emerge from 2003 onwards. A significant development was the integration of Social Network Analysis (SNA) within the learning analytics toolkit. The work of Aviv, De Laat and their colleagues was explicitly situated within the constructivist paradigm that considers knowledge to be constructed through social negotiation (Aviv et al., 2003; De Laat et al., 2006). Their use of SNA, a method developed in the social sciences, allowed them to carry out detailed investigations of networks made up of 'actors' and the relations between them. SNA refers to actors with a relationship between them as 'tied', and these ties can be classified as strong or weak, depending on their frequency, quality or importance (Granovetter, 1973). In the context of learning, social network analysis can be used to investigate and promote collaborative

and cooperative connections between learners, tutors and resources, helping them to extend and develop their capabilities (De Laat et al., 2007; Haythornthwaite, 2006; Haythornthwaite and de Laat, 2010).

Although SNA had strong roots in the learning sciences, it was several years before pedagogic theory had any widespread impact on the literature of learning analytics. Analytic tools were often presented as pedagogy neutral. For example, *GISMO*, a student-monitoring tool, took into account social, cognitive and behavioural aspects of learning. Although its graphical representations allowed teachers to explore these factors, they were not designed to support any specific approach to teaching and learning (Mazza and Milani, 2004). *CourseVis* was also pedagogy neutral, employing LMS data to help instructors understand what was happening in online classes and to identify individuals in need of extra support (Mazza and Dimitrova, 2007).

It was not until 2008 that pedagogic theory started to emerge more strongly in the literature, as an approach to analytics focused on understanding and optimising learning began to crystallise. In part, this was due to the strong pedagogic grounding provided by social network analysts such as Dawson (Dawson, 2008; Dawson and McWilliam, 2008; Dawson et al., 2008). Their social-constructivist view that the process of learning is facilitated through individual participation in social interactions drew on the work of major educational theorists including Dewey (1938) and Vygotsky (1978). Vygotsky's exploration of how knowledge moves between social and individual realms also informed work on collaborative knowledge construction (Suthers et al., 2008) and these groups of researchers shared an interest in the work of Lave and Wenger on situated learning and communities of practice (Lave and Wenger, 1991; Wenger, 1998).

6 Emergence of political and economic drivers

By 2007, learning analytics researchers had begun to address both educational and technological challenges. At this point, development of the field began to increase in pace as it was presented with a new set of challenges, and associated new funding streams.

The *Educause Review* (Campbell et al., 2007) presented a bleak view of the US education system trailing other developed countries, with some graduates lacking even basic competencies. They proposed that 'academic analytics is emerging as a new tool that can address what seem like intractable challenges'. In a separate paper published in the same year, Campbell and Oblinger (2007) set out a definition of academic analytics. This linked the technological, 'Academic analytics marries large datasets with statistical techniques and predictive modeling to improve decision making', with the educational, 'academic analytics has the potential to improve teaching, learning, and student success', in the context of the political, 'by 2020 the overall portion of the U.S. workforce with a college degree will be lower than it was in 2000'.

7 A rapidly expanding field

The influence of political drivers on the field of analytics, together with the growing maturity of the field of EDM – which held its first international conference in Montreal in 2008 – were associated with a split between analytics and EDM. As a result, the

literature of the two diverged and the key EDM references identified by Romero and Ventura (2007) were displaced in the analytics literature by generic references to overviews of the EDM field (Romero and Ventura, 2007; Baker and Yacef, 2009). Despite the division between these fields of enquiry, the future lines for research proposed within these reviews of EDM remain important within analytics research: extending the focus beyond North America, Western Europe and Australia/New Zealand, making mining tools easier for educators to use, standardising methods and data across systems, integrating tools within e-learning environments and developing education-specific mining techniques.

Researchers were already working on tools that responded to these challenges. *Contextualised Attention Metadata (CAM)* addressed the problem of collecting and combining data from different tools by providing a method of collecting metadata from office tools, web browsers, multi-media players and computer-mediated communication, and bringing this together in an attention repository in order to build a rich source of information about user attention (Wolpers et al., 2007).

Increasingly, the analytics tools used to address these challenges were shaped by pedagogical understandings and designed to support learning and teaching. *LOCO-Analyst* provided feedback focused on the quality of the learning process and was explicitly linked by its designers to a requirement for education technology researchers to switch away from technologically driven research towards leveraging technology to meet human needs (Jovanović et al., 2008). The *SMILI☺Open Learner Modelling Framework* was used to support reflection by providing a method for describing, analyzing and designing open learner models (Bull and Kay, 2007). Social network analysis became increasingly influential (De Laat et al., 2007; Borgatti et al., 2009) and the *Social Networks Adapting Pedagogical Practice (SNAPP)* tool was developed to aid analysis of interaction patterns on courses, supporting a focus on areas such as learner isolation, creativity and community formation (Dawson et al., 2010).

Interest was also growing in tools that would allow users to visualise large datasets. *Honeycomb* supported the visualisation of networks including millions of connections (van Ham et al., 2009). The open-source tool *Gephi* supported filtering, clustering, navigation and manipulation of network data (Bastian et al., 2009). A third tool, *sense.us*, supported asynchronous collaboration, including graphical annotation and view sharing, across a variety of types of visualisation (Heer et al., 2009). Other work on visualisation focused on how visual cues could be used to support learning by, for example, increasing student motivation to work with non-mandatory content (Brusilovsky et al., 2009).

One specific tool, *Signals*, developed at Purdue University, became a flagship for academic analytics and is also cited as an example of 'action analytics' that lead to useful outcomes and 'nudge analytics' that prompt individuals to take action (Norris et al., 2009; Arnold, 2010; Carmean and Mizzi, 2010). The *Signals* project mines large datasets and applies statistical tests in order to predict, while courses are in progress, which students are in danger of falling behind. The aim is to produce actionable intelligence, guiding students to appropriate resources and explaining how to use them. A traffic-signal status display shows students whether things are going well (green), or whether they have been classed as high risk (red) or moderate risk (amber). Reported results appear promising; students in the experimental groups sought help earlier than those in the control group, and a pilot group obtained 12% more B/C grades and 14% less D/F grades than a control group (Arnold, 2010).

With tools growing more powerful and their reach increasing, concerns about ethics and privacy began to surface. Should students be told that their activity is being tracked? How much information should be provided to students, faculty, parents, issuers of scholarships and others? How should faculty members react? Do students have an obligation to seek assistance? Although these questions remain pertinent and have not yet been worked through in detail, Campbell (2007) took the initiative in this difficult area, not only by raising these issues but also by proposing a means of addressing them using a framework based on definitions, values, principles and loyalties that could help to locate areas of misunderstanding.

8 Learning analytics emerges as a separate field

In 2010, the field of analytics began to split once again, with learning analytics gradually breaking away from academic analytics. Siemens produced an early definition in an influential blog post

> Learning analytics is the use of intelligent data, learner-produced data, and
> analysis models to discover information and social connections, and to predict
> and advise on learning (Siemens, 2010).

This version was refined following international discussion between researchers, which led to the introduction of the definition cited in the Introduction above. The learning analytics community coalesced around the first international Conference on Learning Analytics and Knowledge, held in Banff, which was followed later that year by the formation of the Society for Learning Analytics Research.

The emergence of learning analytics as a field in its own right meant that there were now separate groupings focusing on each of the challenges driving analytics research.

- Educational data mining focused on the technical challenge: *How can we extract value from these big sets of learning-related data*?

- Learning analytics focused on the educational challenge: *How can we optimise opportunities for online learning*?

- Academic analytics focused on the political/economic challenge: *How can we substantially improve learning opportunities and educational results at national or international levels*?

Overlaps between the three groupings remain, but there have been several attempts to disambiguate the fields. Long and Siemens (2011) focused on current and future meanings, distinguishing between learning analytics – which benefit learners and faculty and are focused at the level of courses and department – and academic analytics – which benefit funders, administrators and marketing at institutional level; funders and administrators at regional level; and governments and education authorities at (inter)national level. Baepler and Murdoch (2010) examined the distinctions between data mining, academic analytics and audit of institutional systems. *Educause*, which has developed its definitions of analytics over several years, has taken a longer-term view, setting out a wider landscape of terminology and highlighting the varied definitions that have emerged over the last decade (van Barneveld et al., 2012).

The development of learning analytics was boosted by their inclusion in the 2011 *NCM Horizon Report* (Johnson et al., 2011). This report was one of a series focused on emerging technologies and their potential impact on and use in teaching, learning and creative enquiry. It identified learning analytics as a technology to watch. The 2012 *NCM Horizon Report* (Johnson et al., 2012) also included learning analytics, judging them to be two to three years from widespread adoption, as well as a subset of the field, social learning analytics.

These analytics are strongly grounded in learning theory and focus attention on elements of learning that are relevant when learning in a participatory online culture (Ferguson and Buckingham Shum, 2012). Approaches to analytics that can be classified in this way include intrinsically social forms of analytic: social network analytics and discourse analytics (De Liddo et al., 2011; Ferguson and Buckingham Shum, 2011). Discourse analytics are a relatively recent addition to the learning analytics toolset, but they draw on extensive previous work in such areas as exploratory dialogue (Mercer and Wegerif, 1999; Mercer, 2000), latent semantic analysis (Landauer et al., 1998) and computer-supported argumentation (Thomason and Rider, 2008).

The term 'social learning analytics' also covers 'socialised' approaches, which can readily be applied in social settings. These include content analytics – recommender systems and automated methods of examining, indexing and filtering online media assets in order to guide learners through the ocean of available resources (Drachsler et al., 2010; Verbert et al., 2011). These analytics take on a social aspect when they draw upon tags, ratings and metadata supplied by learners (see for example, Clow and Makriyannis, 2011). Disposition analytics focus on the experience, motivation and intelligences that influence responses to learning opportunities (Deakin Crick et al., 2004), and are socialised when the emphasis is on the learner engaged in a mentoring or learning relationship.

The development of social learning analytics represents a move away from data-driven investigation towards research more strongly grounded in the learning sciences and, increasingly, dealing with the complexities of lifelong learning that takes place in a variety of contexts. Building up an holistic picture of student progress and taking sentiment into account in order to enable 'computer-based systems to interact with students in emotionally supportive ways' is now seen as a real possibility (Blikstein, 2011, p.110). New tools such as the *GRAPPLE Visualisation Infrastructure Service (GVIS)* do not deal with just one VLE, but can extract data from different parts of a learner's Personal Learning Environment (PLE) and employ these data to support metacognitive skills such as self-reflection (Mazzola and Mazza, 2011).

9 Future challenges

9.1 Challenge 1: build strong connections with the learning sciences

Work focused on cognition, metacognition and pedagogy is under-represented in the key references identified here. Understanding and optimising learning requires a good understanding of how learning takes place, how it can be supported, and the importance of factors such as identity, reputation and affect. As learning analytics emerge from the wide fields of analytics and data mining, disambiguating themselves from academic analytics and EDM, researchers will need to build strong connections with the learning

sciences. This has the potential to be a two-way process, with learning analytics helping to form the basis for good learning design, effective pedagogy and increasing student self-awareness.

9.2 Challenge 2: develop methods of working with a wide range of datasets in order to optimise learning environments

Understanding and optimising the environments in which learning occurs introduces a second challenge. Increasingly, learners will be looking for support from learning analytics outside the VLE or LMS, whilst engaged in lifelong learning in open, informal or blended settings. This will require a shift towards more challenging datasets and combinations of datasets, including mobile data, biometric data and mood data. In order to solve the problems faced by learners in different environments, researchers will need to investigate what those problems are and what success looks like from the perspective of learners.

9.3 Challenge 3: focus on the perspectives of learners

A focus on the perspectives of learners will be essential to the development of analytics related to their needs, rather than to the needs of institutions. Such a perspective has the potential to extend criteria for learning success beyond grades and persistence to include motivation, confidence, enjoyment, satisfaction and meeting career goals. It could also realign work on grading and marking, moving it away from summative assessment that looks back at what learners have achieved, towards formative assessment that helps them to develop. To achieve this will require methods of reporting on and visualising analytics that are personalised, that can be easily understood by learners and that are clearly linked with ways of improving and optimising their learning. In many cases, the analytics process will need to be transparent, enabling learners to respond with feedback that can be used to refine the analytics, and enabling them to see how their data are being used.

9.4 Challenge 4: develop and apply a clear set of ethical guidelines

Meeting these challenges will require decisions regarding the ownership and stewardship of data. The key reference points within the field do not make it clear what rights learners have in relation to their data, or the extent to which they have a responsibility to act on the recommendations supplied by learning analytics. There is no agreed method for researchers to obtain informed and ongoing consent to the use of data, and there are no standard procedures allowing learners to opt out or to have their analytic record cleared. As yet, these issues have been identified as problems, but no detailed ethical framework has been developed for learning analytics. This is a pressing need for the field, and each researcher could play a part here by including a clear section on ethics within their papers and publications.

10 Conclusion

During the last decade, learning analytics has emerged as a significant area of research into technology-enhanced learning. A review of key reference points for the field shows

that a combination of the availability of big datasets, the emergence of online learning on a large scale, and political concerns about educational standards has prompted the development of this field. Learning analytics are distinguished by their concern for providing value to learners, whether in formal, informal or blended settings. They are employed to understand and optimise both learning and the environments within which it takes place. Although this is a new area of research, it draws on extensive work in related areas, and has already developed a range of tools and methods that offer exciting potential. This review has identified four significant challenges that this field must now address: integrating experience from the learning sciences, working with a wider range of datasets, engaging with learner perspectives and developing a set of ethical guidelines. Elsewhere in this issue, Chatti and his colleagues propose a reference model for learning analytics that will support communication between researchers as they seek to address not only these four challenges but also others that will arise as understanding of the technical and pedagogical issues surrounding learning analytics evolves (Chatti et al., this issue).

References

Arnold, K.E. **(2010) 'Signals: applying academic analytics',** *Educause Quarterly***, Vol. 33, No. 1, p.10.**

Aviv, R., Erlich, Z., Ravid, G. and Geva, A. (2003) 'Network analysis of knowledge construction in asynchronous learning networks', *Journal of Asynchronous Learning Networks*, Vol. 7, No. 3, pp.1–23.

Baepler, P. and Murdoch, C.J. (2010) 'Academic analytics and data mining in higher education', *International Journal for the Scholarship of Teaching and Learning***, Vol. 4, No. 2, pp.1–9.**

Baker, R.S.J.D. **and Yacef, K. (2009) 'The state of educational data mining in 2009: a review and future visions',** *Journal of Educational Data Mining***, Vol. 1, No. 1, pp.3–17.**

Bastian, M., Heymann, S. and Jacomy, M. (2009) 'Gephi: an open source software for exploring and manipulating networks', *Paper Presented at the International AAAI Conference on Weblogs and Social Media*.

Berners-Lee, T., Hendler, J. and Lassila, O. (2001) *The Semantic Web*, Scientific American, May 2001, pp.35–43.

Blikstein, P. (2011) 'Using learning analytics to assess students' behavior in open-ended programming tasks', *Paper Presented at the 1st International Conference on Learning Analytics and Knowledge (LAK11)***, Banff, Canada (27 Feb – 1 March).**

Borgatti, S.P., Mehra, A., Brass, D.J. and Labianca, G. (2009) 'Network analysis in the social sciences', *Science*, Vol. 323, No. 5916, pp.892–895.

Britain, S. and Liber, O. (2004) *A Framework for the Pedagogical Evaluation of eLearning Environments*, JISC.

Brusilovsky, P., Sosnovsky, S. and Yudelson, M. (2009) 'Addictive links: the motivational value of adaptive link annotation', *New Review of Hypermedia and Multimedia*, Vol. 15, No. 1, pp.97–118.

Bull, S. and Kay, J. (2007) 'Student models that invite the learner in: the SMILI:-) open learner modelling framework', *International Journal of Artificial Intelligence in Education*, Vol. 17, No. 2.

Campbell, J.P. (2007) *Utilizing Student Data within the Course Management System to Determine Undergraduate Student Academic Success: An Exploratory Study*, PhD, Purdue University.

*Campbell, J.P.***, DeBlois, P.B. and Oblinger, D.G. (2007) 'Academic analytics: a new tool for a new era',** *Educause Review***, Vol. 42(4 July/August), pp.40–57.**

Campbell, J.P. and Oblinger, D.G. (2007) *Academic Analytics*, Educause.

Carmean, C. and Mizzi, P. (2010) 'The case for nudge analytics', *Educause Quarterly*, Vol. 33, No. 4.

Chatti, M.A., Dyckhoff, A.L., Schroeder, U. and Thüs, H. (this issue) 'Learning analytics: a review of the state of the art and future challenges', *International Journal of Technology Enhanced Learning*.

Clow, D. and Makriyannis, E. (2011, 28 Feb – 01 Mar 2011) 'iSpot analysed: participatory learning and reputation', paper presented at *The LAK11: 1st International Conference on Learning Analytics and Knowledge (27 February – 1 March)*, Banff, Canada.

Dawson, S. (2008) 'A study of the relationship between student social networks and sense of community', *Educational Technology and Society*, Vol. 11, No. 3, pp.224–238.

Dawson, S. (2009) 'Seeing the learning community: an exploration of the development of a resource for monitoring online student networking', *British Journal of Educational Technology*, Vol. 41, No. 5, pp.736–752.

Dawson, S., Bakharia, A. and Heathcote, E. (2010) SNAPP: Realising the affordances of real-time SNA within networked learning environments. Paper presented at *The 7th International Conference on Networked Learning*, Aalborg, Denmark (3 – 4 May).

Dawson, S. and McWilliam, E. (2008) *Investigating the Application of IT Generated Data as an Indicator of Learning and Teaching Performance,* Australian Learning and Teaching Council, Canberra.

Dawson, S., McWIlliam, E. and Tan, J.P.-L. (2008) 'Teaching smarter: how mining ICT data can inform and improve learning and teaching practice', paper presented at *The Ascilite 2008*, Melbourne, Australia (30 November – 3 December).

De Laat, M., Lally, V., Lipponen, L. and Simons, R.-J. (2006) 'Analysing student engage-ment with learning and tutoring activities in networked learning communities: a multi-method approach', *International Journal of Web Based Communities*, Vol. 2, No. 4, pp.394–412.

De Laat, M., Lally, V., Lipponen, L. and Simons, R.-J. (2007) 'Investigating patterns of interaction in networked learning and computer-supported collaborative learning: a role for social network analysis', *International Journal of Computer Supported Collaborative Learning*, Vol. 2, pp.87–103.

De Liddo, A., Buckingham Shum, S., Quinto, I., Bachler, M. and Cannavacciuolo, L. (2011) 'Discourse-centric learning analytics', Paper presented at *The LAK11: 1st International Conference on Learning Analytics and Knowledge*, 27 Feb – 1 March.

Deakin Crick, R., Broadfoot, P. and Claxton, G. (2004) 'Developing an effective lifelong learning inventory: the ELLI project', *Assessment in Education: Principles, Policy & Practice*, Vol. 11, No. 3, pp.247–272.

Dewey, J. (1938) *Experience and Education*, Macmillan, London & New York.

Dillenbourg, P. (Ed.) (1999) *Collaborative Learning: Cognitive and Computational Approaches*, Elsevier Science, Oxford.

Drachsler, H., Bogers, T., Vuorikari, R., Verbert, K., Duval, E., Manouselis, N., et al. (2010) 'Issues and considerations regarding sharable data sets for recommender systems in technology enhanced learning', *Procedia Computer*, Vol 1, pp.2849–2858.

Dringus, L.P. and Ellis, T. (2005) 'Using data mining as a strategy for assessing asynchronous discussion forums', *Computers & Education*, Vol. 45, pp.141–160.

EU Expert Group (2010) *Assessing Europe's University-Based Research*, European Commission, Brussels, Belgium.

Ferguson, R. and Buckingham Shum, S. (2011) 'Learning analytics to identify exploratory dialogue within synchronous text chat', paper presented at *The LAK11: 1st International Conference on Learning Analytics and Knowledge*, 27 Feb – 1 March.

Ferguson, R. and Buckingham Shum, S. (2012) 'Social learning analytics: five approaches', paper presented at *The LAK12: 2nd International Conference on Learning Analytics and Knowledge*, 30 April – 2 May.

Granovetter, M.S. (1973) 'The strength of weak ties', *The American Journal of Sociology*, **Vol. 78, No. 6, pp.1360–1380.**

Haythornthwaite, C. (2006) 'Facilitating collaboration in online learning', *Journal of Asynchronous Learning Networks*, Vol. 10, No. 1.

Haythornthwaite, C. and **de Laat, M.** (2010, 3–4 May) 'Social networks and learning networks: using social network perspectives to understand social learning', paper presented at *The 7th International Conference on Networked Learning*, Aalborg, Denmark (3–4 May).

Heer, J., Viégas, F.B. and Wattenberg, M. (2009) 'Voyagers and voyeurs: supporting asynchronous collaborative visualization', *Communications of the ACM*, April, pp.87–97.

Johnson, L., Adams, S. and Cummins, M. (2012) *The NMC Horizon Report: 2012 Higher Education Edition*, The New Media Consortium, Austin, Texas.

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

Jovanović, J., Gašević, D., Brooks, C., Devedžic, V., Hatala, M., Eap, T., et al. (2008) 'LOCO-analyst: semantic web technologies in learning content usage analysis', *International Journal of Continuing Engineering Education and Life Long Learning*, Vol. 18, No. 1, pp.54–76.

Landauer, T.K., Foltz, P.W. and Laham, D. (1998) 'An introduction to latent semantic analysis', *Discourse Process*, Vol. 25, No. 2–3, pp.259–284.

Lave, J. and *Wenger, E.* **(1991)** *Situated Learning: Legitimate Peripheral Participation*, **Cambridge University Press, Cambridge.**

Long, P. **and** *Siemens, G.* **(2011) 'Penetrating the fog: analytics in learning and education',** *EDUCAUSE Review*, **Vol. 46, No. 5, pp.31–40.**

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011) *Big Data: The Next Frontier for Innovation, Competition and Productivity*, McKinsey Global Institute, May.

Mazza, R. and Dimitrova, V. (2004) 'Visualising student tracking data to support instructors in web-based distance education', paper presented at *The WWW Alt. '04: 13th International World Wide Web Conference on Alternate Track*, New York, USA (17–22 May).

Mazza, R. and Dimitrova, V. (2007) 'CourseVis: a graphical student monitoring tool for supporting instructors in webbased distance courses', *International Journal of Human-Computer Studies*, Vol. 65, No 2, pp.125–139.

Mazza, R. and Milani, C. (2004) 'GISMO: a graphical interactive student monitoring tool for course management systems', paper presented at *The T.E.L.'04 Technology Enhanced Learning'04 International Conference*, Milan, Italy (18–19 November).

Mazzola, L. and **Mazza, R.** (2011) 'Visualizing learner models through data aggregation: a test case', paper presented at *The Red-Conference, Rethinking Education in the Knowledge Society*.

McIntosh, N.E. (1979) 'Barriers to implementing research in higher education', *Studies in Higher Education*, Vol. 4, No. 1, pp.77–86.

Mercer, N. (2000) *Words and Minds: How We Use Language To Think Together*, Routledge, London.

Mercer, N. and Wegerif, R. (1999) 'Is "exploratory talk" productive talk?', in Light, P. and Littleton, K. (Eds): *Learning with Computers: Analysing Productive Interaction*, Routledge, London and New York, pp.79–101.

Mitchell, J. and Costello, S. (2000) *International e-VET Market Research Report: A Report on International Market Research for Australian VET Online Products and Services*, John Mitchell & Associates and Education Image, Sydney, Australia.

Norris, D., Baer, L., Leonard, J., Pugliese, L. and Lefrere, P. (2008) 'Action analytics: measuring and improving performance that matters in higher education', *EDUCAUSE Review*, Vol. 43, No. 1.

Norris, D., Baer, L. and Offerman, M. (2009) 'A national agenda for action analytics', paper presented at *the National Symposium on Action Analytics* (21–23 September, 2009).

Rheingold, H. (1993) *The Virtual Community: Homesteading on the Electronic Frontier*, 1st ed.

Romero, C. and Ventura, S. (2007) 'Educational data mining: a survey from 1995 to 2005', *Expert Systems with Applications*, Vol. 33, No. 1, pp.135–146.

Romero, C., Ventura, S. and García, E. (2008) 'Data mining in course management systems: moodle case study and tutorial', *Computers & Education*, Vol. 51, No. 1, pp.368–384.

Siemens, G. (2010) *What Are Learning Analytics?* Available online at: http://www. elearnspace.org/blog/2010/08/25/what-are-learning-analytics/ (accessed on 24 September 2010).

Suthers, D.D., Ravi, V., Medina, R., Joseph, S. and Dwyer, N. (2008) 'Beyond threaded discussion: representational guidance in asynchronous collaborative learning environments', *Computers & Education*, Vol. 50, pp.1103–1127.

Thomason, N. and Rider, Y. (2008) 'Cognitive and pedagogical benefits of argument mapping: L.A.M.P. guides the way to better thinking', in Okada, A. Buckingham Shum, S. and Sherborne, T. (Eds): *Knowledge Cartography: Software Tools and Mapping Techniques*, Springer, London.

Tinto, V. (1997) 'Colleges as communities: taking research on student persistence seriously', *The Review of Higher Education*, Vol. 21, No. 2, pp.167–177.

van Barneveld, A., **Arnold, K.E.** and **Campbell, J.P.** (2012) *Analytics in Higher Education: Establishing a Common Language (ELI Paper 1)*, EDUCAUSE.

van Ham, F., Schulz, H.-J. and Dimicco, J.M. (2009) 'Honeycomb: visual analysis of large scale social networks', *Ifip International Federation For Information Processing*, pp.429–442.

Verbert, K., Drachsler, H., Manouselis, N., Wolpers, M., Vuorikari, R. and Duval, E. (2011) 'Dataset-driven research for improving recommender systems for learning', paper presented at *The LAK11: 1st International Conference on Learning Analytics and Knowledge*, 27 February – 1 March, Banff, Canada.

Vygotsky, L.S. (1978) in Cole, M., John-Steiner, V., Scriber, S. and Souberman, E. *Mind in Society: The Development of Higher Psychological Processes***, Harvard University Press, Cambridge, MA.**

Wenger, E. **(1998)** *Communities of Practice: Learning, Meaning and Identity***, Cambridge University Press, Cambridge.**

*Wolpers, M.***, Najjar, J., Verbert, K. and Duval, E. (2007) 'Tracking actual usage: the attention metadata approach',** *Journal of Educational Technology and Society***, Vol. 10, No. 3, pp.106–121.**

Zaïane, O.R. (2001) 'Web usage mining for a better web-based learning environment', paper presented at *The 4th IASTED International Conference on Advanced Technology for Education (CATE'01)*, 27–28 June, Banff, Canada.