Learning Analytics:
Definitions, Processes and Potential

Tanya Elias
January, 2011
Learning Analytics: The Definitions, the Processes, and the Potential

Learning is a product of interaction. Depending on the epistemology underlying the learning design, learners might interact with instructors and tutors, with content and/or with other people. Many educators expend enormous amounts of effort to designing their learning to maximize the value of those interactions. Regardless of the approach taken, a series of questions consistently arises: How effective is the course? Is it meeting the needs of the students? How can the needs of learners be better supported? What interactions are effective? How can they be further improved?

Traditional approaches to answering these questions have involved student evaluation, the analysis of grades and attrition rates, and instructor perceptions most often gathered at the end of a course. Consequently the evaluation and analysis of learning has suffered from: a limited quantity of data busy students and instructors are willing to share at the end of a course; the limited quality of this self-reported, retrospective data; and a significant delay (normally at least one semester) between the events being reported and the implementation of an intervention. As an increasingly large number of educational resources move online, however, an unprecedented amount of data surrounding these interactions is becoming available. This is particularly true with respect to distance education in which a much higher proportion of interactions are computer-mediated. For example, the amount of time reading content online can be easily captured by an LMS/CMS. When, why and with whom learners are connecting is also logged in discussion forums and social networking sites.

Recently, interest in how this data can be used to improve teaching and learning has also seen unprecedented growth and the emergence of the field of learning analytics. In other fields, analytics tools already enable the statistical evaluation of rich data sources and the identification
of patterns within the data. These patterns are then used to better predict future events and make informed decisions aimed at improving outcomes (Educause, 2010). This paper reviews the literature related to this emerging field and seeks to define learning analytics, its processes, and its potential to advance teaching and learning in online education.

Learning Analytics and Related Concepts Defined

Learning analytics is an emerging field in which sophisticated analytic tools are used to improve learning and education. It draws from, and is closely tied to, a series of other fields of study including business intelligence, web analytics, academic analytics, educational data mining, and action analytics.

*Business Intelligence* is a well-established process in the business world whereby decision makers integrate strategic thinking with information technology to be able to synthesize “vast amounts of data into powerful, decision making capabilities” (Baker, 2007, p.2). *Web analytics*, is defined as “the collection, analysis and reporting of Web site usage by visitors and customers of a web site” in order to “better understand the effectiveness of online initiatives and other changes to the web site in an objective, scientific way through experimentation, testing, and measurement” (McFadden, 2005). A particularly powerful way to gather business intelligence, it involves the compilation of data from hundreds, thousands, and even millions of users during which trends are noted, hypotheses are formed, and alterations to the website based on those hypotheses can be implemented and tested (Rogers, MacEwan and Pond, 2010). It also demonstrates the use of increasingly complex computer-mediated data-tracking, capture and modelling to meet the current needs and predict the future needs of their customers (Cho et al., 2002; Mobasher et al., 2000; Wang & Ren, 2009). Analytics software might, for example, evaluate data mined from purchasing records to suggest products that might interest customers or
allow a search engine to target ads based on an individual’s location and demographic data (Educause, 2010). Through the application of these processes, businesses have been able to “provide the user with a more personalised, relevant and timely experience and therefore, provide the company with a better bottom line” (Dawson et al., 2010).

Goldstein and Katz (2005) coined the term *academic analytics* to describe the application of the principles and tools of business intelligence to academia. Their goal was to study the technological and managerial factors that impact how institutions gather, analyze, and use data. Campbell and Oblinger (2007) used a narrower definition of the term academic analytics in that they opted to study issues directly related to “one of higher education’s most important challenges: student success.” They identified student retention and graduation rates as the two most common measurements (p.1). Unlike educational data mining, which seeks to search for and identify patterns in data, “academic analytics marries large data sets with statistical techniques and predictive modeling to improve decision making” (ibid, p.3).

Norris et al. (2008) further emphasized the importance of using educational data to act in a forward-thinking manner in what he referred to as *action analytics*. Action analytics included deploying academic analytics “to produce actionable intelligence, service-oriented architectures, mash-ups of information/content and services, proven models of course/curriculum reinvention, and changes in faculty practice that improve performance and reduce costs.” Similarly, Arnold (2010) spoke of analytics as a tool whereby institutions would:

have the potential to create actionable intelligence on student performance, based on data captured from a variety of systems. The goal is simple—improve student success, however it might be defined at the institutional level. The process of producing analytics frequently challenges established institutional processes (of data ownership, for example), and initial analytics efforts often lead to additional questions, analysis, and implementation challenges.
Norris et al. (2008) identified a number of colleges in the process of deploying academic analytics including Baylor University, University of Alabama, Sinclair Community College, Northern Arizona University and Purdue University which are changing decision-making, planning, and resource allocation processes related to resource utilization, student retention and student success at a grassroots level.

Dawson et al. (2010), however, complained that:

> While the Horizon report recognises the growing need for more HE institutions to provide more detailed and sophisticated reportage, the report falls short in discussing the advantages and opportunities available for teaching and learning in accessing institutional captured data.... Access to these data has traditionally been removed from the learning context and has only recently begun to expand into the scholarship of teaching and learning. However, further expansion is necessary. (p. 124)

Learning analytics seems aimed at addressing this concern. Next Generation: Learning Challenges (n.d.) identified goal of this emerging field as the ability to “scale the real-time use of learning analytics by students, instructors, and academics advisors to improve student success.” Thus, the focus appears to be on the selection, capture and processing of data that will be helpful for students and instructors at the course or individual level. Moreover, learning analytics is focused on building systems able to adjust content, levels of support and other personalized services by capturing, reporting, processing and acting on data on an ongoing basis in a way that minimizes the time delay between the capture and use of data. Thus, in contrast to current evaluation processes which use the results from one semester to inform improvements in the next, learning analytics seeks to combine historical and current user data to predict what services specific users may find useful now.

Dawson (2010) cited the following example.

> Although it is now accepted that a student’s social network is central for facilitating the learning process, there has been limited investigation of how networks are
developed, composed, maintained and abandoned. However, we are now better placed than our predecessors to use digital technologies for the purpose of making learner networking visible.... If teachers are enabled to ‘see’ those who are network-poor earlier in their candidature, it becomes possible for them to make timely and strategic interventions to address this issue. (p.738)

Thus, learning analytics seeks to capitalize on the modelling capacity of analytics: to predict behaviour, act on predictions, and then feed those results back into the process in order to improve the predictions over time (Eckerson, 2006) as it relates to teaching and learning practices. Currently however, the built-in student tracking functionality in most CMS/LMS are far from satisfactory (Hijon and Carlos, 2006) and do not offer sufficient learning activity reports for instructors to effectively tailor learning plan that meet the needs of their students (Zhang et al., 2007). Thus, the study and advancement of learning analytics involves: (1) the development of new processes and tools aimed at improving learning and teaching for individual students and instructors, and (2) the integration of these tools and processes into the practice of teaching and learning.
Learning Analytics Processes

Many representations of the analytical process have been developed over time in a variety of disciplines. Despite their diverse origins, they have much in common and are helpful in identifying a set of processes essential for the implementation of learning analytics.

*Knowledge continuum.* In his development of an actionable knowledge conceptual framework for business, Baker (2007) used a much older “knowledge continuum” as a starting point (He traces it back to the 1800s). Raw data is at the bottom of the continuum. It consists of characters, symbols and other input that, on its own, is meaningless. As meaning is attached to this data, it becomes information. Information is capable of answering the questions who, what, when and where. Through analysis and synthesis that information becomes knowledge capable of answering the questions why and how. Finally, that knowledge is transformed into wisdom through its application. Baker suggested that predictive analytics and the development of actionable knowledge corresponded with the transformation of knowledge to wisdom. The knowledge continuum highlights that it is in the processing and use of data that it is transformed into something meaningful.

Large stores of data already exist at most colleges and universities, and computer-mediated distance education courses are increasingly creating trails of student data. By analyzing this data, analytics applications have the potential to provide a predictive view of upcoming
challenges, both for the institution and for students. This data could be used to make data-driven decisions able to support optimal use of both economic and pedagogical resources while offering a structure for improved educational outcomes (Educause, 2010).

Despite the depth and range of data available and its ability to inform a diversity of end-users, to date there has been limited application of this data within higher education (Dawson et al., 2010). Thus, despite the presence of data, educators often lack the specific information they need to identify important performance issues. Moreover, academic culture favours analysis over action; institutions have placed a high degree of importance on their reputations rather than on improving academic performance of their students (Norris, 2008). Thus in the majority of institutions, the development of actionable knowledge related to learning has been stalled at the data level with the collection of a large amount of data in a meaningless form.

**Web analytics objectives.** As if in response to the institutions stalled in the data level of the knowledge continuum, Rogers, MacEwan and Pond (2008) explain, “there are so many metrics that *could* be tracked that it is absolutely essential for stakeholders to identify what types of outcomes they desire from users” (p. 233). Hendricks, Plantz and Pritchard (2008) identified four objectives essential to the effective use of web analytics in education: define the goals or objectives, measure the outputs and outcomes, use the resulting data to make improvements, and share the data for the benefit of others. By defining goals and using those goals to determine what data to capture, educators run less risk of “drowning in data” (Snibbe, 2006). Furthermore they highlight that these are not steps in a process, but rather opportunities to ask probing questions to enable success: What do we want to achieve? Are we measuring what we should be measuring? How will this information be used? How can we create “innovative metrics and mashups to illuminate deeper outcomes?” (p. 245).
Mazza and Dimitrova (2004), for example, used the goal of helping teachers become aware of social behaviour and cognitive aspects of remote learners to guide the development of CourseVis, a system of visualization and graphical representation of web log data generated by the WebCT CMS. Similarly, Purdue University’s Signals project emerged from a desire to provide students with real-time feedback on their course progress in an intuitive format (Arnold, 2010). Although the outcomes of these two projects vary significantly, they demonstrate the need to identify a goal at the beginning of the project and carefully select the data to be used accordingly. In this way, Rogers, MacEwan and Pond (2010) suggest “researchers and practitioners in distance education may in fact be uniquely positioned to take the use of analytics data in design process and strategic decision-making to a new level” (p.245).

The five steps of analytics. Campbell and Oblinger (2008) described academic analytics as an “engine to make decisions or guide actions” that consists of five steps: capture, report, predict, act, and refine. Like the knowledge continuum, these steps begin with the capture of meaningless data which is then reported as information, to enable predictions based on knowledge and wise action. The addition of the final step refine recognizes analytics as a “self-improvement project” in which “monitoring the impact of the project is a continual effort, and statistical models should be updated on a regular basis” (p.8).

Despite the recognition of the importance of ongoing improvement of the system in learning analytics, the literature related to this process is scarce. Outside of education, search engines, recommenders and ratings systems evident on many commercial sites are excellent examples of how data gathered during an analytics cycle can be used to further refine offerings for users. The integration of these types personalization in education has the potential to advance the development of personalized learning environments.
Collective Application Model. In their work on the design of collective applications, Dron and Anderson (2009) present a model that is also useful in defining the processes of learning analytics. Their model consisted of five layers divided into three cyclical phases. In their explanation of the model they stated:

If we do not re-present actions to the crowd through an interface that affects similar actions, it is just data mining for some other purpose. This is not a knowledge discovery cycle. (p.369)

Their model also emphasizes the cyclical nature of analytical processes and the on-going need to refine and improve the system through successive phases of gathering, processing and presenting information. In the wider sphere of learning analytics, these phases can be equated to gathering, processing and application. Gathering involves data selection and capture. Processing includes the aggregation and reporting of information and making predictions based on that information. Finally, application involves the use, refinement and sharing of knowledge in attempts to improve the system.

By comparing and combining the models and frameworks described above, seven related processes of learning analytics emerge: Select, Capture, Aggregate & Report, Predict, Use, Refine and Share (see Table 1).
Learning Analytics: Definitions, Processes and Potential

Table 1: Comparison of Analytics frameworks and models

<table>
<thead>
<tr>
<th>Knowledge Continuum</th>
<th>Five Steps of Analytics</th>
<th>Web Analytics Objectives</th>
<th>Collective Applications Model</th>
<th>Processes of Learning Analytics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>Capture</td>
<td>Define goals</td>
<td>Select</td>
<td>Select</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Measure</td>
<td>Capture</td>
<td>Capture</td>
</tr>
<tr>
<td>Information</td>
<td>Report</td>
<td>Aggregate</td>
<td>Aggregate &amp; Report</td>
<td></td>
</tr>
<tr>
<td>Knowledge</td>
<td>Predict</td>
<td>Process</td>
<td>Predict</td>
<td></td>
</tr>
<tr>
<td>Wisdom</td>
<td>Act</td>
<td>Use</td>
<td>Display</td>
<td>Use</td>
</tr>
<tr>
<td></td>
<td>Refine</td>
<td></td>
<td></td>
<td>Refine</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Share</td>
</tr>
</tbody>
</table>

Learning Analytics Tools and Resources

Learning analytics is almost always associated with powerful computers and sophisticated programming capable of processing vast quantities of data. Dron and Anderson pointed out, however, that the analytical process is a “single amalgam of human and machine processing which is instantiated through an interface that both drives and is driven by the whole system, human and Machine” (p. 369). Similarly, Hackman and Woolley (in press) identified that analytics was cognitive, technical and social in nature. These findings support the earlier work of Sharif (1993) who identified technology as a combination of both the hardware and the knowledge-skills-abilities required to effectively use it: technoware, humanware, infoware, and orgaware. By combining this idea with Bogers and Daguere’s (2002) conception of technology as a body of knowledge Baker

Figure 3: Technological resources (Sharif, 1993 as cited in Baker, 2007)
(2007) depicted technology resources as seen in Figure 3. More simply put, learning analytics consist of computers, people, theory and organizations.

**Computers.** When one thinks of technology resources, computers and software are often what come to mind. Sharif (1993) refers to these as technoware. As previously mentioned, higher education institutions, particularly those involved in distance education, already gather an enormous amount of student data. Dawson et al. (2010) noted:

> The information on student behaviour captured by the LMS has been so rarely interrogated and adopted beyond basic load and tool usage. The quantity and diversity of data available regarding student online learning behaviour, interactions with peers and teaching staff and access to other institutional ICT systems (student services, library, etc.) for example, affords an opportunity for integrating automated student learning support and services. (p.121)

Thus, clearly many data gathering tools are already in place. The challenge with respect to data-gathering hardware and software is the integration of these diverse data sources. Open-architecture solutions capable of scraping data, information, and context from administrative and academic systems and from structured and unstructured data, information, and context contained in assessment solutions are therefore required (Norris et al., 2008). If LMS/CMS data were correlated with additional information gathered in other systems, a richer picture of student learning experience, instructor adoption and institutional usage could be generated. It could in fact be possible to track individual activity throughout the entire student life cycle – from initial admission, through course progression and finally graduation and employment transitions (Dawson, 2010).

Once the data is gathered, a variety of reporting and predictive tools are required to process the information. These tools include data visualization, decision trees, neural networks, regression analysis, machine learning, and artificial intelligence (Corbit, 2003). Shemwell (2005) suggested that visual displays are critical to decision making as humans can process an
incredible amount of information if it is packaged and presented correctly. Digital dashboards are critical data visualization tools. Common presentation techniques include charts, graphs, gauges, dials, and maps (Baker, 2007). With respect to education, Mazza and Dimitrova (2007) insisted that the graphical representation of these data could aid instructor interpretation and evaluation of pedagogy in action. Numerous authors have found that meaningful information can be extracted from CMS/LMS and be made available to instructors and students in the form of dashboard-like interfaces that incorporate predictive models (Campbell & Oblinger, 2007; Morris et al., 2005; Wang & Newlin, 2002; Arnold, 2010).

Figure 4: Purdue University's SIGNALS dashboard. Retrieved from http://news.uns.purdue.edu/images/+2009/signals-screen.jpg

<table>
<thead>
<tr>
<th>Name</th>
<th>Total Views</th>
<th>Sessions</th>
<th>Online Time</th>
<th>Viewed Resources</th>
<th>Initial Threads</th>
<th>Total Posts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average</td>
<td>35750</td>
<td>00:59</td>
<td>1:15:56</td>
<td>00:56</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>506</td>
<td>70</td>
<td>00:44</td>
<td>00:11:54</td>
<td>47</td>
<td></td>
<td></td>
</tr>
<tr>
<td>203</td>
<td>40</td>
<td>00:04:52</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>304</td>
<td>74</td>
<td>12:54:34</td>
<td>36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>54</td>
<td>11:59:09</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>52</td>
<td>00:21:48</td>
<td>37</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>200</td>
<td>67</td>
<td>10:26:45</td>
<td>32</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>260</td>
<td>51</td>
<td>14:42:18</td>
<td>85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>740</td>
<td>105</td>
<td>10:54:17</td>
<td>47</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>972</td>
<td>72</td>
<td>05:57:26</td>
<td>49</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 5: Moodle Student Activity Report, Zhang (2010).
While the use of dashboard technology is growing in popularity, there are challenges to providing the right information to the right people right away that remain. Few (2007) noted that, although visually appealing, many dashboard technologies lack the ability to provide truly useful information. Ensuring the timeliness, accuracy and usefulness of the source data supporting what information is displayed on the dashboard is a critical challenge that can render a decision support tool useless (Baker 2007). Moreover, MacFayden and Dawson (2010) found that for the purposes of monitoring student activity and achievement, predictive models must be developed at the course level and that evaluative and data visualization resources must be highly customizable to cater to instructor differences for adopting LMS tools and their overarching pedagogical intent.

Another visualization tool gaining popularity in learning analytics facilitates social network analysis (SNA). SNA draws on various concepts from graph theory and structural theory to evaluate network properties such as density, centrality, connectivity, betweenness and degrees. These measures provide a framework for interpreting and developing an understanding of the observed patterns of exchanges that occur between social actors. In online learning, student data can be gathered about various types of communication including chat logs, discussion forum postings, blog posts and comments. The potential richness of these sites for social network data mining provides (Dawson, 2010).

Social Network Analysis tools essentially automated the process of extraction, collation, evaluation and visualisation of student network data, quickly presenting network

Figure 6: Social Networks Adapting Pedagogical Practice visualization retrieved from http://research.uow.edu.au/learningnetworks/seeing/snapp/index.html
Increasingly, educators are turning to computers to not only gather data and process information, but also to apply knowledge. Zhang (2007) stated:

In addition to collecting and presenting student activity data, we can proactively provide feedback to students or the instructor. Moodog tracks the Moodle logs, and when certain conditions are met, Moodog automatically sends an email to students to remind them to download or view a resource. (p. 4417)

It is likely that there will be an increasing level of dependence on computers to implement interventions in learning analytics.

*Theory.* Computer hardware and software are only useful if they are based on sound knowledge. In the case of learning analytics, theory includes both analytics-related knowledge and good practice accumulated in other fields. Offering the exact and time-sensitive recommendations that have been widely applied in many e-commerce systems, for example, depends on the use of recommendation methods based on different theories such as collaborative filtering algorithm, bayesian network, association rule mining, clustering, hurting graph, knowledge-based recommendation, etc. and the use of collaborative filtering algorithms (Cho, 2009).

Accumulated knowledge related to areas such as learning theory, sound pedagogical practices, building knowledge communities, student motivation, perseverance and motivation, and student retention is also essential in learning analytics. MacFayden and Dawson (2010) found however:

Very little research exists, and no guidance is available for educators, to indicate which (if any) of the captured tracking variables may be pedagogically meaningful – that is to say, which of the many available data points are indicative of student participation in educationally purposeful activity that may contribute to their learning and achievement in a course. (p. 590)

At Purdue, students, instructors and administrators all mentioned significant variation in the intervention models facilitated by analytics and the need for best practices to be established.
Tone, frequency, method of communication, and intervention schedules are only a few of the other areas needing further research (Arnold, 2010).

Moreover, the performance target for graduating students is shifting rapidly and today’s measures such as grades, transcripts and resumes are already proving to be insufficient measures of performance and potential. As the pace of knowledge creation, sharing, and use continues to accelerate, it becomes increasingly important for education to focus on the development of complex, sophisticated, and applied skills in teamwork, critical thinking, communication, and aggressive perpetual learning rather than the simple development of subject-specific competencies (Norris et al., 2008). Teaching these skills will involve ongoing pedagogical change based on ongoing teaching and learning research.

**People.** Though modern technology applications are hardware and software that replaces human effort with machine effort, there are still a significant number of aspects of the system that require knowledge, skills, and abilities of humans to ensure effective operation (Ahasan and Imbeu 2003). Astin (1993), for example, demonstrated that robust and diverse peer networks lay an important role in student persistence and overall academic success. However, not all attempts at developing highly engaging social interactions are effective (MacFayden and Dawson, 2010). Although teaching staff can use the network visualization tools to reflect on the design, timing and evaluation of implemented learning activities designed to achieve a set pedagogical agenda such as the generation of a learning community (Dawson, 2008), developing effective interventions remains highly dependent on human cognitive problem-solving and decision-making skills.

**Organizations.** The most often overlooked technology resources is the relational, or social, capital required in the field of analytics: Who will be involved in an analytics project? How
well are their decisions and actions supported? How well can they communicate and interact with one another? Woolley et al. (2010) identified the existence of collective intelligence which “is not strongly correlated with the average or maximum individual intelligence of group members but is correlated with the average social sensitivity of group members, the equality in distribution of conversational turn-taking, and the proportion of females in the group” (p. 686). When undertaking learning analytics projects it will therefore be essential to support the development of their social skills and enhance the communication tools of team members.

Moreover, regardless how good the work of the analytics team, to truly put apply knowledge, institutions need committed leaders who will consciously build organizational capacity to measure and improve performance and to change organizational culture and behaviour. Achieving faculty buy-in on performance measurement and improvement and the ongoing use of analytics solutions in their instructional practice is paramount to the institution’s ability to build and sustain a culture of evidence-based action. For-profit higher education institutions have focused their operational policies and practices on “actions that work” for adult learners; the University of Phoenix and Capella University consistently make extensive use of artificial intelligence and predictive modeling in marketing, recruitment, and retention and have shaped their cultures around performance. Over time, sage leadership and commitment can develop the technology, information, and analytics capabilities of colleges and universities sufficiently to create more evidence-based and action-oriented behaviour and culture (Norris et al., 2008).
Learning Analytics Moving Forward

Put together, learning analytics uses of four types of technology resources to complete ongoing pattern of three-phase cycles aimed at the continual improvement of learning and teaching as illustrate in Figure 7.

Figure 7: Learning analytics continuous improvement cycle

Through this process high-quality data, current pedagogical theory and human innovation can work together to better meet the needs of learners, educators and administrators. Norris et al. (2008) envisioned analytics capable of better assessing student competencies and a system that, through the use of individualized planning, advising, and best practices from cradle to career, is able to align interventions to learner need and fully maximize learner success. The tools to realize this vision do exist, thus the challenge for the field of learning analytics will be to
facilitate and support the change required at all levels of education in order to fully realize this potential.
References


