**Failing Students need Big Data and Learning Analytics:  Hype or Reality?**

Dr Ray Stoneham
Department of Computing and Information Systems, University of Greenwich

Big data and learning analytics, like MOOCs, have suffered from hype and inflated expectations, as depicted in the Gartner Hype Cycle for Education (Maguire, 2014).  It is time to evaluate the promise of big data and learning analytics before they reach their peak of expectation and prior to the trough of disillusionment, with the hope of assisting them up the slope of enlightenment to the plateau of productivity. This opinion piece attempts to do this.

'Big data' is difficult to define. For some, it means 'high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making' (Gartner, 2014), whilst for others it is '[a] new attitude by businesses, non-profits, government agencies, and individuals that combining data from multiple sources could lead to better decisions" (Bechtold 2014). There seems to be more common agreement about the definition of learning analytics as 'the interpretation of a wide range of data produced by and gathered on behalf of students in order to assess academic progress, predict future performance, and spot potential issues' (Johnson et al, 2011). Whatever the definitions, however, it is important to differentiate between hype and reality.

Universities collect vast amounts of data about their students and store and analyse this for a variety of purposes; many are using, or proposing to use, the techniques of big data and learning analytics to do more.  This data includes explicit factual information obtained as part of the application and registration process, such as gender, ethnicity, UCAS points score, home postcode and prior educational achievements, and explicit ongoing data accumulations, such as attendance, assessment results and even students' use of counselling services (Parr 2013). Universities also collect the implicit digital data trails of students' online activity, such as their use of the VLE, intranet and online library systems, and perhaps also their use of email and the internet. Many universities are now planning to make better use of this mass of small and big data. One of the driving forces behind this is to predict the success, or otherwise, of individual students with the hope that analysis of this data early in the academic year will allow corrective action to be taken before it is too late (de Quincey and Stoneham, 2014; Dimopoulos et al, 2013; Dyckhoff et al, 2012).  But, as in 'The Minority Report' (Dick, 1956), where prediction of homicidal actions is used in an attempt to prevent murders, there are dangers posed by these automated predictive methods (Dowland, 2014).

Analysis of explicit and implicit student data of students in the Department of Computing and Information Systems at the University of Greenwich has been used for several years to help improve retention of students and some moderate success has been achieved that could be attributed to this approach.  A student dashboard presents a range of data about each student, including attendance, intranet usage, network usage, uploads, feedback, grades, personal tutorial records, assessment offences (if any) and enquiries.  Administrators, managers and tutors make use of this when making decisions and providing support. No attempt has been made to generate a single 'student engagement percentage' for each student and there are no automatically-generated emails when thresholds are met. Instead,

the system provides the data in a format that helps staff to identify problems and take appropriate action, including recording any actions taken for others to see.

Other universities have implemented different systems. Nottingham Trent University (Foster and Gascoigne, 2014), for example, has a system that allocates a single 'Student Engagement' percentage for all students, but, as details of how it is calculated are not publicly available, it is impossible to evaluate. Some other systems are described in a Jisc report (Sclater, 2014), but evaluative testing of all systems is currently quite limited. It is important that the algorithm used has different adjustable weightings of the components and that the requirements of different universities, faculties and departments are taken into account. The one-size-fits-all approach does not, for example, allow for the probability that the use of the library by students of English or History is much more significant to their progress than its use by Digital Media or Architecture students. Similarly, to make a single Student Engagement percentage act as a trigger for automated emails needs careful consideration: emails sent to students with poor engagement and with known extenuating circumstances could have unfortunate consequences.

The data gathered also needs to be interpreted with care, to avoid potential bias and oversimplification. Stephen Fry's comment (Fry, 2010), that "Education is the sum of what students teach each other between lectures and seminars", illustrates the fact that students learn in different ways (Garcia et al, 2007). Automated pattern recognition and data mining can hold students prisoner to their past choices and performance by assuming that they cannot change their behaviour and that their personal circumstances have not changed. Ineffective and misdirected interventions resulting from faulty learning diagnoses might result in resentment and broken trust.

There is also the problem of students attempting to 'game' the system (Bollier, 2010). Students have been known to sign in for absent students to boost their attendance figures and they could manipulate the data in other ways, for example, by borrowing books for no purpose or by clicking on intranet pages to boost the number of their hits. In an increasingly high-stakes litigious university environment, such student outcries as "Why did you not help me when you knew I was struggling?" or "How dare you interfere with my learning!" might become common. Meanwhile, data about staff are also recorded in the system and there are questions about how these will be used by the institution.

In conclusion, it is clear that students learn in different ways and that one size certainly does not fit all. Failing students can be identified and helped by the use of big data and learning analytics, but identification of 'failing students' is difficult and the use of automated response systems problematic. Ethical issues need to be discussed and agreed at institutional level before any systems are implemented, access to the data and systems should be carefully controlled and the systems must be understood and transparent. Using big data and learning analytics does entail risk, but it is to be hoped that progress up the slope of enlightenment to the plateau of productivity will bring benefits for all students and institutions.

## Reference list

Bechtold, D. (2014) *What is Big Data?* Available at: http://www.oldcolony.us/wp-content/uploads/2014/11/whatisbigdata-DKB-v2.pdf (Accessed: 6 June 2014).

Bollier, D. (2010) *The Promise and Peril of Big Data.* Available at: http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_o f_Big_Data.pdf (Accessed: 1 July 2015).

de Quincey, E. and Stoneham, R. (2014) 'Student Engagement: An evaluation of the effectiveness of explicit and implicit Learning Analytics.' Presentation at ePIC 2014 - 12th International ePortfolio & Identity Conference, Greenwich, UK.

Dick, Philip K. (1956) *The Minority Report.* First published in Fantastic Universe, January 1956, and reprinted in: *The Variable Man and other stories,* 1957. New York: Ace Books.

Dimopoulos, I., Petropoulou, O., Boloudakis, M. and Retalis, S. (2103) 'Using Learning Analytics in Moodle for assessing students' performance.' 2nd Moodle Research Conference, Sousse, Tunisia. Available at: http://research.moodle.net/mod/data/view.php?d=1&rid=152 (Accessed: 21 October 2014).

Dowland, P. (2014) *Collecting data on students: is it useful to know which books they've read?* Available at: http://www.theguardian.com/higher-education-network/blog/2014/apr/29/universities-collect-data-on-students-learning (Accessed: 29 April 2014).

Dyckhoff, A. L., Zielke, D., Bültmann, M., Chatti, M. A. and Schroeder, U. (2012) 'Design and Implementation of a Learning Analytics Toolkit for Teachers.' Educational Technology & Society, 15 (3), 58–76.

Foster, E. and Gascoigne, G. (2014) *NTU student dashboard: Learning analytics to improve retention.* Available at: http://www.ntu.ac.uk/webevents/news_events/document_uploads/efye_manual/7_1_foster_g ascoigne.pptx (Accessed: 2 June 2014).

Fry, S. (2010) *The Fry Chronicles.* London: Michael Joseph.

Garcia, P., Amandi, A., Schiaffino, S. and Campo, M. (2007) 'Evaluating Bayesian networks' precision for detecting students' learning styles.' Computers & Education, 49, (3) 794–808. Available at: http://www.sciencedirect.com/science/article/pii/S0360131505001788 (Accessed: 30 June 2015).

Gartner IT Glossary (2015) *Big Data.* Available at: http://www.gartner.com/it-glossary/big-data (Accessed: 30 June 2015).

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

Maguire, D. (2014) *The 'hype cycle' of Moocs and other big ideas.* Available at: http://www.timeshighereducation.co.uk/comment/opinion/opinion-the-hype-cycle-of-moocs-and-other-big-ideas/2010206.article (Accessed: 10 January, 2014).

Parr, C. (2013) *Universities mine institutional data in search of gold.* Available at: http://www.timeshighereducation.co.uk/news/universities-mine-institutional-data-in-search-of-gold/2002053.article (Accessed: 1 March 2013).

Sclater, N. (2014) *Learning analytics: The current state of play in UK higher and further education.* Available at: http://repository.jisc.ac.uk/5657/1/Learning_analytics_report.pdf (Accessed: 30 June 2015).