25 Can Learning really be Managed – Big Data and the Future of the Learning Management Systems

David EVANS
JME Associates, UK

This paper begins where the Netties 2017 Conference in Athens concluded, with the work of Greek philosophers. I'll then look at how theories of learning have developed more recently, along with new ideas about styles of learning before moving to look at attempts, using computers to capture, measure, manage and control learning. This has culminated in the current era of big data, where huge quantities of information about individuals are available to various private and public agencies which can be manipulated in a bewildering variety of ways by those who know what they're doing. The implications are wide reaching and frightening. The current concerns about the nefarious use of data in political persuasion are highly visible - more wide-ranging concerns about the use of data in health, medicine and finance are also developing. How will these debates play out, and what will happen when big data really hits education?

Many of our current ideas about learning began in Greece around 2500 years ago. Although the invention of writing was obviously important, Socrates, the first Greek philosopher wrote nothing down. In his discussions of the mythical origins of writing in Ancient Egypt, Socrates (through Plato) is concerned that it will introduce forgetfulness, because people will not practice using their memory. Real learning comes through discussion, argument and self-reflection, or what we would now call rationalism. It was left to Plato's student Aristotle to suggest more engagement with the real world. Aristotle was the first empiricist - using his senses to look for knowledge and learning by attempting to categorise and measure things in the real world. One can assume however that none of them would have been great fans of learning management systems.

With the ancient Greeks also began the distinction between what we might call liberal and vocational education – learning to think and learning practical skills. Although liberal education in philosophy and the arts was considered superior, vocational education did take place, and what we now call the apprenticeship system developed alongside the school system as alternative methods of learning. Indeed there are plays, 5th century comedies by Pherecrates and Nichomachus about the training of cooks. These two parallel approaches to education and training continued into the Roman empire.

Compared with the timespan between Ancient Greece and the present day, the development of theories of learning, attempting to measure learning and the capability of doing so are of much more recent origin. It wasn't until the development of psychology as a discipline in the late 19th century that empirical approaches to the learning process itself began to develop. Until that time, learning had been assumed to take place using the methods barely changed since Plato's academy at the top end - the elite universities - and mass schooling methods which originated in Prussia from the 18th century onwards for the masses. Within the limits of their time, both these systems seemed to work, and were widely emulated. There was no need to be further concerned about or to try to measure what, if anything, people were really learning.

However, on the other side of the world, Ancient China had introduced the first standardized tests, the imperial examination to select candidates for governmental positions in about 600 AD. Examinations for entry in to the British Civil Service, modelled on the Chinese system were introduced in 1806, and there were similar schemes in other countries. Written examinations didn’t begin to be used in British universities until the 1850s, although oral exams had been used since the middle ages. In the early 20th century, the mass conscription associated with the First World Wars led to an expansion of standardised testing, the US Army making extensive use of the recently invented Stanford-Binet IQ system for this purpose.

When the psychological study of learning really began in the early 20th century, it led to a further dichotomy - between the behaviourists, who saw it as a largely mechanical process, the same in pigeons and human beings - simply learning to do things - and the cognitivists, who saw it as slightly more complicated than that, an internal process that occurred within the minds of learners who each had to construct their own knowledge. At the same time, in the wake of rapid expansion in science and all forms of knowledge and the development of mass schooling, the need for learning to be more efficient, wide ranging and cost effective became paramount.

There has also been much debate in recent decades about whether different people learn in different ways - issues of 'learning styles' (visual, auditory or kinaesthetic; theoretical, practical or pragmatic), sometimes connected to what Howard Gardner calls 'multiple intelligences'. These approaches have both zealous advocates and unimpressed critics, so I won't dwell on them here, except to observe that in constructing any form of learning experience, it's usually a good idea to recognise that different people learn in different ways and where possible, appreciate choices in both how to undertake and if necessary, how to demonstrate their learning

At this stage, we need to recognise that the distinction between behaviourism and other theories of learning was more than just a scientific debate. It had implications for how, if at all, anyone could really measure learning. For learning to have taken place, the individual who has learned should be able to demonstrate that learning. For the most part, this involved exams or tests - but in the last resort, these always imply a behaviourist approach, as they merely demonstrate that someone can do something, or regurgitate something in a controlled environment within a time limit. They don't really show that he or she actually knows anything - or will remember it tomorrow, next week or in a few years' time. Even now, attempts to measure or capture this kind of information are rare – Kirkpatrick's four stages of evaluation usually stop at stage 1.

In the aftermath of the second world war, theories of learning developed in parallel with the explosion of computing power. Although mechanical teaching machines had already been around since about 1920, the first systems for computer aided instruction, Plato (programmed logic for automatic teaching operations at the University of Illinois) wasn't developed until the early 1960s. This was followed by a plethora of Computer Based Learning or Training systems, particularly in the military and specialist areas such as aviation - and there were also early experiments in Computer Managed Learning. CAI, CBT and CML continued through a series of largely incompatible systems, moving onto PCs during the 1980s before the invention of the World Wide Web in the 1989 led to the prospect of managing learning much on a much larger scale. This was a time when wild predictions were being made about all education being online within five years. As John Chambers of Cisco systems said in 1999 in an often quoted remark, "education over the Internet is going to be so big it is going to make e-mail usage look like a rounding error".

The Aviation Industry Computer Based Training Committee's work in the 1980s led to the US military Advanced Learning Initiative SCORM (Sharable Content Object Reference Model) framework around 2000. Both were designed to promote interoperability of learning content, which would otherwise be locked into incompatible formats with a correspondingly short shelf-life. This was when training management and record keeping systems, or what are misleadingly called Learning Management Systems first developed. In industry, the LMS is increasingly seen as a legal requirement to ensure compliance training has been undertaken, and it's also used for talent management and succession planning. In universities and schools, the emphasis is a bit more on helping students to learn, although whether students see it in that way is open to debate.

The problem with SCORM is that in the last resort, all it really measures is the amount of time a learner had spent nominally using, or logged in to a piece of learning content, and / or the score he or she had achieved in a specific test. Some other bits of other information can be tracked and recorded though most LMS - but only activity that occurs with the purview of the LMS. Any other learning can't easily be included. The same issues also exist with face to face education - a student can sleep through a lecture, but will still be marked as having been present.

This was at a time when increasing attention was being paid to the fact that most learning appears to happen outside of any of the normal contexts within which it had traditionally been defined. The 70:20:10 model developed by Charles Jennings suggests that only 10% of learning actually occurs through formal courses. 20% occurs thorough other structured interventions, but 70% occurs informally, on the job, through interaction with fellow students or whatever. In other words, it isn't and can't be tracked or measured through SCORM

Jennings talks of learning as a continuous process, artificially broken into discrete events. Learning management is therefore really learning process management – but in many organisations, around 2005, learning effectively became the LMS, and LMS vendors made a lot of money. A lot of these organisations are now trying to extricate themselves from bad LMS deals. An LMS is useful for process automation, training administration, regulatory and compliance testing. It's less good for actually helping learners. The question is should we extend or limit its role – should we aim for total learning management, or simply better learning analytics?

In the early 21st century, Jennings suggests, the 2,000 year old model of the school was overturned, through the development of blended learning, and what's sometimes called the flipped classroom. At the same time, work has been changing rapidly – jobs requiring decision making over 40% in USA, increasing every year. The important thing is speed to competence – chunks or learning, not long theoretical courses, involving the growing use of video, social media, social learning and mobile devices.

In 2011, ADL recognized the need for a software specification that tracks learning experiences that occur outside of a LMS and a web browser. This led to the development of the Experience API, often known by its project name of Tin Can, which potentially allows for the capture of much more information on human performance, along with associated learning content or contextual information. It also enables dynamic tracking of activities from any platform or software system—from the traditional LMS to mobile devices, simulations, wearables and more, including the Internet of Things.

xAPI-enabled learning activities generate statements, or records of e-learning in the form of "I did this" or "Actor / verb / object". These statements are then stored in a Learning Record Store, which can be part of a learning management system (LMS), or a separate database, such as the open source Learning Locker. LRSs can communicate learner data with other systems, such as LMSs, sensor-enabled devices, mobile technology, and other LRSs. Importantly, xAPI statements are capable of being sent to multiple LRSs at once. With traditional LMSs, a learner's data stays with the organization that administers the LMS. With an LRS, the data can follow the learner - from job to job or from school to school. Individual learners can also have their own LRSs, or Personal Data Lockers, in which they store learning data for their own personal records.

LRSs offer the ability to create in-depth e-learning analytics because of the large amounts of much more wide-ranging, potentially “big” learning data they record and store. How can this emerging big data help learning? It supposedly helps shopping with Amazon suggesting that if we enjoyed one book, we might like another book on a similar topic, based on data collected from many thousands of customers. Until recently, such techniques could not have been used to improve the learning experience as individual educational institutions simply don’t have enough useful student data, but MOOCS are different? By observing thousands of students' patterns of behaviour online, can we help them learn more effectively?

Ken Cukier (in Learning with Big Data) describes an example from the University of Arizona: through monitoring keystrokes and hesitation patterns during a course, they noticed that in lesson 7, a lot of students return to lesson 2 slides to check on basic maths principles. This kind of thing enables intervention, courses redesign etc, to reduce dropout – and provides an economic incentive for universities and colleges to invest. He also cites Duolingo (where more people are learning Gaelic than the current number of native speakers), which earns revenue by selling correct translations to industry as a by-product of free online learning, as does Google translate.

Big data, whether from MOOCS or elsewhere is not generated by a research question – it’s simply there, as a by-product of interactions with the system. This is not the usual order of things in science, where a hypothesis, or research question would normally determine the data gathered. It’s more similar to using official statistics, collected for administrative purposes, but which can be used for social research.

With large-scale learning platforms, such as Coursera, EdX and FutureLearn, learning data does not have to remain merely local. If the learning design (the planning and management of learning activities) can be used by other teachers all collecting the same data from their students, it can be widely tested, with many independent peers able to review, advise and redesign for a better outcome. If many teachers run the same design through their local virtual learning environment (VLE), with students using the same digital tools to collect their performance data, you have effectively crowdsourced local data.

However, there are several problems. One is to involve real subject matter specialists in the process. Often, the field is dominated by technology professionals who are not educators and don’t do much teaching online – other than teaching about how to teach online. In the same way, much discussion about learning is dominated by learning professionals – real learners often aren’t so bothered about their styles or psychological processes. If more lecturers could be helped to collaborate and generate their own large-scale data collection and analysis, perhaps educational big data could really make a difference.

Big data offers new ways to socially sort with increasing precision. By combining huge data sets, a lot can be learned through ‘algorithmic profiling’, but it raises concerns about how little people know about how their data is collected. As we've seen with Facebook, data brokers can take vast sets of data and perform correlations, selling them on to other agencies for nefarious purposes. Low educational attainment is, for example, might be associated with greater propensity to gambling, higher spending on junk food, likelihood of being influenced by fake news and so on. There have been many cases of people whose credit ratings have been downgraded simply because others who shopped where they shop had a poor repayment history. ‘Creditworthiness by association’ could equally become educational attainment level by association.

There are numerous examples of data breaches in recent years. These can lead to identity theft, blackmail, reputation damage and distress. Most examples are of credit data - but health data and education data also vulnerable.

In addition, blacklisting and watch-lists sometimes simply identify people incorrectly. It has been found that being wrongfully identified in this case can negatively affect employment, ability to travel – and in some cases lead to wrongful detention and deportation.

We need to learn from these issues. While there are a range of individuals and groups developing ideas about how data harms can be prevented, there are others working in the opposite direction. Alexander Nix, who is currently suspended from his post as head of Cambridge Analytica, told an undercover BBC reporter that he wanted access to the US market 'while the data laws were still the wild west'. Researchers, civil society organisations, government bodies and activists have all, in different ways, identified the need for greater transparency, accountability, systems of oversight and due process, and the means for citizens to interrogate and intervene in the big data processes that affect them, including those concerned with learning. What is needed is the public pressure and the political will and effort to ensure this happens.