10 „Big Data in Education“

DI Dr. Erwin BRATENGEYER
Danube University Krems, Austria
erwin.bratengeyer@donau-uni.ac.at

10.1 Abstract

Big data has arrived in education. Reformers and data scientists see student data as the key to the future of education. In contrast, critical pedagogues advise against dehumanization through decision making by statistics and data security officials warn of privacy, legal and ethical issues inherent to learning data. Pros and cons of big learning data are exposed.

10.2 Introduction

According to business intelligence analysts big data means big impact. The obvious current excitement surrounding big data has been generated primarily from the web and those leading e-commerce vendors with their product recommender systems. Likewise many areas of science and technology are exploiting the data of a multitude of sensors from any kind of activities and events whether detected from machines, nature, or living beings, whether stemming from the micro or the macro cosmos. The Internet of Things and mobile devices are transforming our society with their impact on business, entertainment, healthcare and politics. But what about education? Did big data arrive at schools and at universities? According to a quick check with Google Trends using the keywords <big data education> big data popped up in education only recently in October 2012, for whatever reason, and is continuing to gain in importance (see Fig. 1).

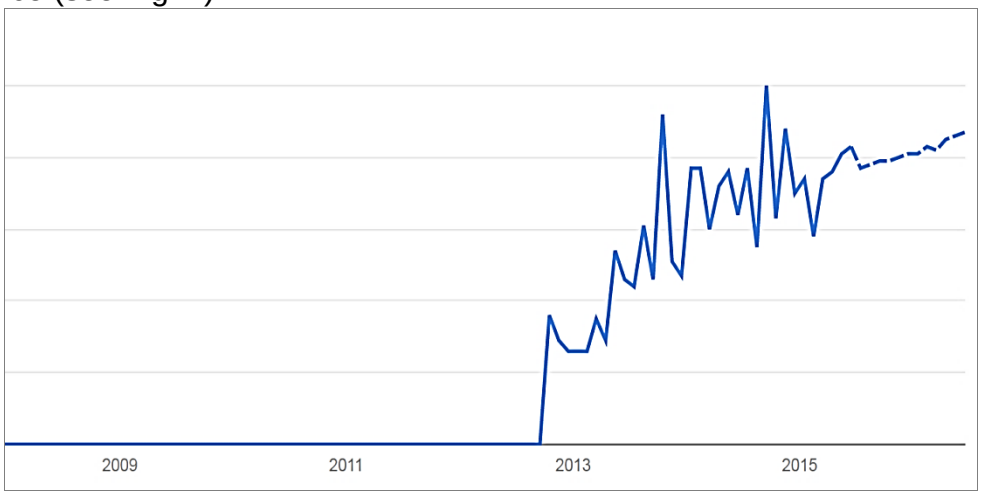

Fig 1: Google Trends result on the keywords <big data education>
Source: https://www.google.com/trends, [30.06.2015]

It seems worth taking a closer look at what big data in education is about. Digging deep into computerized statistical data on learners to make decisions regarding their performance is a trend that excites some and worries others.

10.3 Big Learning Data

Without doubt ICT became increasingly embedded in the educational domain. E-learning has become a cornerstone in higher education, in schools and in the work place. The number of online courses is steadily increasing worldwide. Learners and employees leave tracks while they are interacting with lecturers, trainers and any kind of content material, data about their successes and failures can be captured even in real time. Big data in education is created when learners enroll and engage in learning activities with content which is mainly provided through a learning management system (LMS). Those LMSs generate data from logins, courses, modules, files, assignments, exams, quizzes, grades, learning paths, and different kinds of interactions with peers, tutors, and lecturers. Thanks to a newly introduced e-learning specification, the Experience API (xAPI), even more data is created blowing the limitations of browser-based LMSs. xAPI allows for recording learning experiences from real-world activities, experiential learning, social learning, offline learning, or collaborative learning. Those activity streams contribute considerably to the increase of data records.

Ferreira (2015) divides educational data into five types, one on student identity and four on student activities, in order of how difficult they are to attain: (1) *Identity Data* - who are you? what is your role? (2)*User Interaction Data* - engagement metrics like click rate, page views, (3)*Inferred Content Data* - efficacy data on instructional materials, (4)*System-Wide Data* - attendance**,** grades. (5)*Inferred Student Data* - exactly what concepts does a student know and need? For building recommendation systems *User Interaction Data* and simple rules are not enough. According to Ferreira *Inferred Content* and *Inferred Student Data*, though difficult to obtain, are key to successfully build such adaptive systems.

An increasing number of companies is keen on datafying the learning process claiming to satisfy the needs of learners by improving comprehension and performance as well as the needs of administrators by improving nothing less than the whole education system. Feinlab (2015) posted a list of innovative education analytics applications which helps to illustrate the current market situation and product features. Here are a few examples: *CourseSmart* embeds analytics directly into digital textbooks measuring how students are interacting with their books allegedly allowing for accurately predicting student outcomes. *Declara* developed a dynamic map of how people learn based on years of research across hundreds of subjects taking every interaction an individual has over the network to deliver the right content. *Knewton* personalizes digital courses figuring out what each student knows and how that student learns best, then recommends what to study next.

While companies are pushing for the aggregation of student data into analytics tools to improve their algorithms concerns arise over who should have access to what data, for what purposes. The achievements of data scientists in developing algorithms seem to proceed faster than the competences of other stakeholders on what to do with the outcomes. Certainly the debate over the use of data in educational contexts is only at its beginning.

10.4 Benefits and Challenges

The vast availability of user interaction data is setting the course for unlocking the potential of learning analytics. So the expectations are high. According to Alacron (2014) education reformers see the merging of student data, predictive analytics, processing tools, and technology-based instruction as the key to the future of education and a means to further opportunity and equity in education. Education researchers and companies promise that personalization will bring about a more adaptive, responsive, and efficient school system and ameliorate inequalities in the educational system. Proponents advocate data-driven instruction and educational reform as a low-cost and more effective means to deliver better content to a wider variety of students, which, in turn, is seen as a means to encourage social and economic mobility.
From an institutional point of view the promises point at facilitating both global and community development, help marketing study programs, can improve graduation rate, facilitate early warning systems and at least can optimize the enrollment process. From a learner´s point of view big data can personalize the learning process, can measure performance beyond test scores, can increase interaction via mechanics, can monitor a student´s level of engagement, and can produce adaptive learning plans. “Personalized learning” adjusts instructional content, pace, and complexity to meet an individual learner's needs and objectives.

So much for the benefits. What about the downside? As mentioned above data-driven educational policy is a polarizing topic. Concerns about privacy and security of personal data, and legal and ethical issues are raised unmistakably. Disclosures about global surveillance have raised public interest and consequently gathering data is no longer a niche service. Privacy and security issues have to be treated by transparent policies with increased urgency, however, such policies are not around yet. Legal and ethical questions have to be answered but discordant stakeholders – critical pedagogues, teachers, learners, parents, learning analytic evangelists, administrators, developers, and data security officials get in on the act. Data ownership is a very sensitive issue in learning contexts and this issue isn´t solved yet.

Similarly disturbing is the potentially dehumanizing impact of statistics and probabilistic. Profiling substitutes calculation for human judgment on what should be very sensitive human issues, and thus treats those profiled as objects, as collections of facts, rather than as persons. Relying just on algorithms and numbers when it comes to decisions on personal careers is frightening. The more frightening it is when we forget to keep in mind that the relevance of data is not always to be taken as granted, correlation does not prove causation. The risk of spurious correlations - associations that are statistically robust but happen only by chance - increases with more data. Big data analysis needs to stand on solid ground thus proper skills must be developed. Research in this domain requires increased attention. In short on could say: big data without big analysis is big mistake and without big care is big brother.

Viktor Mayer-Schönberger, Professor of Internet Governance and Regulation at Oxford University and Kenneth Cukier, a prominent commentator on developments in big data elaborately deal with the pros and cons in their publication on learning with big data (Mayer-Schönberger & Cukier, 2014). Among other things they come to a

far-reaching socio-critical conclusion as a possible result of the big data impact on education: “As these changes unfold, we´ll find that many of the tools and institutions we rely on must themselves change. […] This will lead to an unbundling of the educational experience. The monopoly that schools hold today is starting to resemble the monopolies once held by monarchy and Church. It is poised to crumble, as did those other seemingly impregnable institutions when the currents of an earlier information revolution – printing – washed over them.” Will education change – after all?

10.5 References

Alarcon, A., et al., (2014): Data & Civil Rights: Education Primer (October 30, 2014). Data & Civil Rights Conference, Available at http://dx.doi.org/10.2139/ssrn.2542268 [30.06.2015].

Feinlab, D., (2015): The Big Data Landscape Blog. Available at http://www.bigdatalandscape.com/blog/big-data-higher-education-apps-change-everything [30.06.2015].

Ferreira, J., (2015): Big Data in the Education System. Blog. Available at https://www.brainscape.com/blog/2014/11/big-data-education-system/ [30.06.2015].

Mayer-Schönberger, & V., Cukier, K., (2014): Learning with Big Data – The Future of Education. Houghton Mifflin Harcourt.