# Education 2.0: Learning Analytics, Educational Data Mining and Co.

**Tim Jülicher**

**Abstract** Internet devices and digital resources such as MOOCs or social networking services (SNS) have become an essential element of learning environments that provide a wide variety of data. Using educational data mining (EDM) and learning analytics (LA), we can gain detailed insights into students' learning behavior. Predictive analytics enable adaptive learning, databased benchmarking and other novelties. However, in Europe big data and education are not big topics yet; whereas, in the US the discussion about both the potentials and risks of linking and analyzing educational data is gaining momentum. Problems arise from the use of educational apps, classroom management systems and online services that are largely unregulated so far. That is particularly alarming with regard to data protection, IT security, and privacy. Last but not least, the analysis of personal educational data raises ethical and economical questions.

## 1 Digitization in Educational Institutions

School as we know it has barely changed since its invention in the 17th century.[1] Tests on paper, use of blackboards, lessons with textbooks—despite all methodical and pedagogical reforms there has long been no sign that these basic things would ever change. However, classrooms and lecture halls have undergone gradual changes in recent years: Nowadays, smart boards can be found in many schools, lectures come along with complementary online services[2] and students may

[1] See Kerstan (2013), Die ZEIT 25/2013.

[2] A prominent example is the open source learning platform Moodle (www.moodle.org), which is available in more than 220 countries.

T. Jülicher (✉)
Institute for Information, Telecommunication and Media Law (ITM),
University of Münster, Münster, Germany
e-mail: tim.juelicher@uni-muenster.de



T. Hoeren and B. Kolany-Raiser (eds.), *Big Data in Context*,
SpringerBriefs in Law, https://doi.org/10.1007/978-3-319-62461-7_6

participate by using real-time feedback and voting apps.[3] Even most preschool kids are already familiar with tablet computers,[4] smartphones and the like.

Against this background, the whole learning process is expected to undergo significant changes. New technologies—particularly digitization—will most likely transform the educational sector, i.e. schools and universities, considerably.

## 2 The Future of Education—Predictions and Benchmarking

For decades if not centuries, individual learning behavior was measured by exams, grades, credit points, or certificates. Nowadays, there are numerous technologies that allow for far more sophisticated insights of rather informal nature. For example, detailed data cannot only be retrieved for individuals (i.e. a single student), but also for groups (e.g. classes) or specific clusters (e.g. multiple institutions in a school district).

- For instance, when using an e-learning platform, teachers can see how often their latest slides have been downloaded. At the same time, they can retrieve statistics on how long students were logged in and what they were occupied with in the meantime.[5]
- When using massive open online courses (so-called MOOCs) it is possible to analyze each and every participant's clickstream. That allows drawing conclusions on the individual's learning behavior and her/his potential shortcomings.[6]
- Where tablets, smartphones and e-books replace conventional textbooks, we can retrace at which pace a student reads and which passages s/he read to prepare herself/himself for the exam. We can even figure out whether s/he actually fulfilled the compulsory reading task at all.

Those examples show that new technologies and concepts—such as integrated learning (so-called blended learning)[7]—generate huge volumes of data, which might relate to the learning behavior of individual students, the learning progress of

[3]See for instance the German open source project "ARSnova", the app "SMILE—Smartphones in der Lehre" or commercial services like "Top Hat".

[4]In Germany, some schools have "iPad classes", cf. the pilot project "Lernen mit iPads" at a Montessori school in Cologne-Bickendorf, http://monte-koeln.de/wp-content/uploads/2014/07/Konzept_iPad.pdf.

[5]For detailed information see https://docs.moodle.org/28/en/index.php?title=Course_overview_report.

[6]Breslow et al. 2012, Research & Practice in Assessment 8/2013, p 11 et seqq.; Picciano, JALN 3/2012, p 13 et seq.

[7]Blended learning means that education is not restricted to face-to-face classroom situations, but includes instructions and materials by means of digital and online media (i.e. computer-mediated activities).

an entire class or the success of a teaching concept. However, gathering the data is only the first step. The real promises (and challenges) of education 2.0 lie in the second step. That is linking and analyzing the data.

## 3 Educational Data Mining and Learning Analytics

For this purpose, techniques such as educational data mining (EDM) and learning analytics (LA) come into action. The first task is to properly organize all information that has been gathered in the course of a completely digitized learning process. That usually involves two types of information: structured and unstructured data. While structured data includes information such as a learner's IP address or her/his username, unstructured data may relate to various texts from internet forums, video clips, or audio files. Beyond that, there is so-called metadata (e.g. activity data or content-related linkage).

At first, educational data mining allows extracting relevant information, organizing it and putting it into context—regardless of its original function. That is a crucial step in order to process the data for further analytical purposes.[8] In this regard, as a matter of fact, educational data mining resembles what is known as commercial data mining. Commercial data mining describes the systematical processing of large datasets in order to gain new, particularly economical insights.[9] It is often used in financial or industrial contexts.[10]

Subsequently, learning analytics seeks to interpret the collected data and draw conclusions from it.[11] Basically, the underlying idea is to optimize the individual learning process by exploiting the provided raw data. This does not only include a comprehensive visualization and reproduction of *past* learning behavior. It rather aims for predicting *future* learning behavior. This process is called predictive analytics.[12] It allows, for instance, to detect and tackle individual learning deficits at an early stage in order to prepare a student for the next assignment.

[8]Cf. Ebner/Schön 2013, Das Gesammelte interpretieren—Educational Data Mining und Learning Analytics.

[9]Nettleton 2014, Commercial Data Mining, p 181 et seqq.

[10]Reyes, Tech Trends 2015(59), p 75 et seq.

[11]For further information on how EDM and LA act in concert see Siemens and Baker 2012, LAK'12, p 252 et seqq.

[12]Siemens 2010, What Are Learning Analytics? http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/; for further examples of use see Sin/Muthu, ICTACT 2015 (5), p 1036.

## 4 Stakeholder

Who benefits? On one hand, it is students and teachers, of course. Teachers can not only retrieve summarized reports for entire classes but also track the learning behavior of individual students. This allows them to respond at an individual level and with measurements that are tailored to the student's particular needs. Lecturers, too, are able to receive in-depth and real-time feedback regarding their teaching behavior and skills.

In addition, educational data might be useful for scientific or administrative purposes. One might want to evaluate institutions, lecturers, curricula or student profiles, for example. When it comes to structural reforms, educational data might also come handy for political actors.[13]

Last but not least, there are quite a number of economic players. As a matter of fact, it is not only software developers or hardware manufacturers that long to know more about the use of their products. Schoolbook publishers or market actors that provide learning-related services such as private lessons, coaching or retraining are among the interested parties, too. Employment agencies and HR departments are certainly interested in educational data as well.

## 5 Data Sources and Data Protection

While private sector companies have discovered big data as an emerging technology some time ago (think of buzzwords like industry 4.0 or the internet of things), in terms of education big data does not seem worthy of discussion yet. This is quite surprising given the fact that students, teachers, and lecturers generate a considerable amount of data. That, in turn, raises the question: Where does educational data come from?

Since the US educational system proves to be more liberal in implementing new technologies, it provides some ideas. Teachers in the US increasingly rely on so-called classroom management systems (CMS) or mobile apps to organize their classes. However, very few of these applications are actually tested and/or approved by supervisory authorities. Therefore, the use of educational software is hardly regulated. Besides, many apps lack common IT security standards. From a quality point of view,[14] very few apps guarantee sufficient accuracy standards.

Apps are just one source of educational data. Additional information might come from e-learning platforms, laptops and tablets (sometimes provided by schools) or

[13]That is what Reyes calls "educational decision making", see Tech Trends 2015 (59), p 77. See also Williamson, Journal of Education Policy 2016, p 125 et seqq.

[14]Regarding data quality see chapter "Big Data and Data Quality" in this book, p 8 et seqq.

student IDs with RFID functionality.[15] Data from social network services (SNS) might be involved as well.[16] A common problem with these sources relates to unauthorized access. In some cases, third parties are likely to have access to more data than schools and universities have. At least, that is what one provider of adaptive learning systems claims.[17]

After all, it is the student who provides most educational data.[18] That is crucial since s/he is often a minor who is usually protected by specific underage provisions. As s/he (involuntarily) discloses personal and highly sensitive data, privacy activists fear the Orwellian "transparent student".

Against this background, it is astonishing that big data in education is not a controversial topic in Europe yet. Particularly, since European jurisdictions have considerably higher standards regarding privacy and data protection. In Germany, for instance, there are not only supranational (GDPR) and federal provisions (BDSG) but also school-specific regulations at state level (e.g. sections 120–122 SchulG NRW).[19]

## 6 Summary and Challenges

To put it in a nutshell, big data promises revolutionary changes in education. It is true, as Slade and Prinsloo point out, that "[h]igher education cannot afford to not use data".[20] However, the difference between the US educational system and European—particularly German—schools and universities does not only result from a different level of technological implementation but also from unequal privacy legislation standards.

Since educational big data technologies are still in their infancy in Europe, all relevant stakeholders should take the chance to enter into a joint discussion as early as possible. Such a dialogue should focus on a critical reflection of promises and risks. First and foremost, it needs to take into account aspects like privacy, data protection, transparency and individual freedoms. After all, tracking and analyzing

[15]Hill 2014, A Day In The Life Of Data Mined Kid, http://www.marketplace.org/topics/education/learningcurve/day-life-data-mined-kid.

[16]Reyes, Tech Trends 2015 (59), p 78.

[17]That is what Jose Ferreira 2012, CEO of Knewton, claims about his product, https://www.youtube.com/watch?v=Lr7Z7ysDluQ.

[18]However, it should be taken into consideration that teachers and lecturers become transparent, too. That is particularly relevant from a data protection point of view.

[19]In the US there is a wide range of sector-specific privacy legislation to follow. See National Forum on Education Statistics 2016, Forum Guide to Education Data Privacy, p 2 et seqq.; Varella, Rutgers Computer & Technology Law Review 2016 (42), p 96 et seqq.

[20]Slade/Prinsloo 2013, American Behavioral Scientist 57(10), p 1521.

an entire educational career creates unforeseeable implications for both individuals and the society as a whole.[21]

## References

Breslow L, Pritchard DE, DeBoer J, Stump GS, Ho Seaton DT (2012) Studying learning in the worldwide classroom. research into edX's first MOOC. Res Pract Assess 8:13–25

Ebner M, Schön M (2013) Das Gesammelte interpretieren—educational data mining und learning analytics. In: Ebner M, Schön S (eds) Lehrbuch für Lernen und Lehren mit Technologien. Epubli. http://l3t.tugraz.at/index.php/LehrbuchEbner10/article/download/119/117. Accessed 4 Apr 2017

Ferreira J (2012) Knewton—Education Datapalooza. https://www.youtube.com/watch?v=Lr7Z7ysDluQ. Accessed 4 Apr 2017

Hill A (2014) A day in the life of a data mined kid, Marketplace September 15. http://www.marketplace.org/topics/education/learningcurve/day-life-data-mined-kid. Accessed 4 Apr 2017

Kerstan T (2013) Wer hat die Schule erfunden? Die ZEIT 25/2013

Koska C (2015) Zur Idee einer digitalen Bildungsidentität. In: Gapski H (ed) Big data und Medienbildung. Koepad, Düsseldorf/München, pp 81–93

National Forum on Education Statistics (2016) Forum guide to education data privacy. (NFES 2016-096). U.S. Department of Education Washington, https://nces.ed.gov/pubs2016/Privacy_Guide_508_7.6.16.pdf. Accessed 4 Apr 2017

Nettleton D (2014) Commercial data mining: processing, analysis and modeling for predictive analytics projects. Morgan Kaufmann/Elsevier, Amsterdam/Boston

Picciano AG (2012) The evolution of big data and learning analytics in American higher education. J Asynchronous Learn Netw 16(3):9–20

Prinsloo P, Slade S (2013) An evaluation of policy frameworks for adressing ethical considerations in learning analytics. In: LAK'2013 proceedings of the third international conference on learning analytics and knowledge, pp 240–244

Reyes J (2015) The skinny on big data in education: learning analytics simplified. https://nces.ed.gov/pubs2016/Privacy_Guide_508_7.6.16.pdf. Tech Trends 2015 (59):75–79

Siemens G, Baker RS (2012) Learning analytics and educational data mining: towards communication and collaboration. In: Proceedings of the 2nd international conference on learning analytics and knowledge, pp 252–254

Siemens G (2010) What are learning analytics? http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/. Accessed 4 Apr 2017

Sin K, Muthu L (2015) Application of big data in education data mining and learning analytics—a literature review. ICTACT J Soft Comput 5:1035–1049

Slade S, Prinsloo P (2013) Learning analytics: ethical issues and dilemmas. Am Behav Sci 57 (10):1510–1529. doi:10.1177/0002764213479366

Varella L (2016) When it rains, it pours: protecting student data stored in the cloud. Rutgers Comput Technol Law J 2016(42):94–119

Williamson B (2016) Digital education governance: data visualization, predictive analytics, and 'Real-time' policy instruments. J Educ Policy 31(2):123–141

[21]For further information on the ethical implications see Slade/Prinsloo 2013, American Behavioral Scientist 57(10), p 1510 et seqq. and Prinsloo/Slade, LAK 2013, p 240 et seqq.; that particularly affects the "digital educational identity", as some authors point out, see Koska 2015, in: Gapski, Big Data und Medienbildung, p 82 et seqq.

## Author Biography

**Tim Jülicher** B.A., Dipl.-Jur., research associate at the Institute for Information, Telecommunication and Media Law (ITM) at the University of Münster. He holds degrees in law and political sciences from the University of Münster.