Big Data Technologies in K-12 Education

Kira Lowe

Introduction

In a world where data-capturing devices are found everywhere from our home computers and televisions to our mobile phones and watches, many private industries are harnessing the power of abundant, available data to increase revenue and improve customer experience. We are perpetually reminded via highly targeted emails and Facebook ads that our personal data is in constant use by advertisers, but few of us are aware of how and if the public sector is harnessing the power of big data. More specifically, are educational institutions jumping on the big data bandwagon? Are they able to effectively generate and utilize data to predict student achievement, identify potential problems, and improve student outcomes?

Making use of educational data to improve schools is not a new concept. In the 1940s, the government used the earliest computers to develop educational data models (Gunawardena, 2017). By the mid-90s, a key algorithm was developed that allowed for the genesis of a process the authors called "knowledge tracing;" this algorithm was applied to a test preparation tutoring program which allowed for personalization of instruction based on each individual student's mastery of a skill (Corbett & Anderson, 1994). The concept of knowledge tracing is still in use today. When the *No Child Left Behind* legislation was introduced in 2001, effective use of data in strategic planning became an important component of education administration and accountability (Mitchell & Conrad, 2003). Since then, not only has data collection on the performance and perceptions of students and teachers multiplied, but also data collection methods have become increasingly digital.

Foster et. al define big data as "data with high volume, high velocity, and great variety" (Foster, Ghani, Jarmin, Kreuter, & Lane, 2016). Thus far in the twenty-first century, all three of these qualities are becoming increasingly satisfied. In the early 2000s, the advent and proliferation of online course management systems, such as Blackboard, has provided a platform for the digitization of traditionally analog data. Gradebooks that were previously kept on paper can now live online which allows them to be exported for further analysis. Additionally, course management systems include features such as forums and chats that generate new forms of exportable data. Other software developments, like digital textbooks that can track how much time a student spent reading each section of course material, create data on previously inaccessible statistics (Gunawardena, 2017). Even more sophisticated data is currently emerging as biometric measurement of non-cognitive skills becomes a new point of study (Roberts-Mahoney, Means, & Garrison, 2016, Shechtman, DeBarger, Dornsife, Rosier, & Yarnall, 2013).

Despite the increasing volume, velocity, and variety of data available today, simply possessing data does not necessarily translate into improved outcomes. Teachers and school administrators often struggle to manage big data and their changing roles and responsibilities in the educational ecosystem (Mitchell & Conrad, 2003; Roberts-Mahoney et al., 2016). As policies and technologies continue to evolve, data-driven instruction will become progressively present in classrooms. It is imperative that policy makers, administrators, educators, parents, and students understand when and how data-driven instruction can or should be used in their schools, and what implications those uses have.

Turning data into insight

Two disciplinary fields lay the foundation for data scientists to cull information out of the vast amount of data we have access to today: Educational Data Mining (EDM) and Learning Analytics (LA). Liñán and Pérez describe both EDM and LA as being concerned foremost with extracting and analyzing large amounts of data with the goal of improving educational outcomes. They both deal heavily in prediction, clustering, relationship mining, knowledge tracing, and other methods that ultimately intend to distill a large amount of complex information into actionable insights (Liñán & Pérez, 2015). Even though they employ many of the same methodologies and ultimately have similar goals, EDM and LA have some important distinguishing elements that differentiate implications for their use.

Educational Data Mining focuses particularly on automation of both the analysis of data and decision-making based on the knowledge gained from that analysis. It utilizes a combination of artificial intelligence, statistical knowledge, data mining, and machine learning technologies (Liñán & Pérez, 2015). EDM relies heavily on predictive models that collect and synthesize components of a data set in order to make inferences about those components based on all available information (Baker & Inventado, 2014). As Liñán and Pérez suggest, perhaps the best use case for this field of study is in Adaptive Learning Systems, which adopt EDM models to rapidly adjust curriculum to the learner in real time. These systems continuously gather, synthesize, analyze and identify patterns in the data, then predict an outcome, adapt curricula to that prediction, and then repeat the process (Liñán & Pérez, 2015). This operation is unique to every student user and therefore provides a curriculum tailored to that individual. It should come as no surprise, then, that "EDM has strong origins in educational software, student modeling, and predicting course outcomes" (Romero & Ventura, 2013). While a number of tools exist that utilize EDM methods, many of them are not widely available to the public or are too complicated for the average teacher or staff member without a background in data mining to use (Liñán & Pérez, 2015).

Learning Analytics has evolved roughly in tandem with Educational Data Mining, but the two disciplines employ different tools with slightly different purposes. LA lends more of an emphasis to social network analysis and sentiment analysis than EDM does (Liñán & Pérez, 2015). Additionally, whereas automation is a fundamental element of EDM, LA uses automation simply as a tool that allows for the leveraging of human judgement; whereas EDM is rooted in educational software and concerns itself primarily with automated discovery and adaptation of components of a system, LA is more concerned with the system as a whole and gives predictive and interventional information to teachers and, in certain cases, to students as well (Romero & Ventura, 2013; Liñán & Pérez, 2015).

Educational Data Mining and Learning Analytics share a common goal of improving outcomes for students, teachers, and schools, even if the techniques and methodologies differ. Results of both practices can have the effect of increased feedback to teachers and their managers which can be instrumental to strengthening the education system. Additionally, the same types of feedback can be provided to students in order to empower them to set their own goals and giving them greater control and accountability for their actions and behaviors (Roberts-Mahoney et al., 2016).

Developments and opportunities

In both the public and private educational ecosystems, pressure is building for leaders and educators to incorporate the latest technology and tools into instruction. Educational Data Mining has been shown in multiple studies to increase not only the likelihood, but also the visibility of improved learning outcomes; an alluring asset to administrators in an age of accountability (AlShammari, I., Aldhafiri, & Al-Shammari, Z., 2013). In addition to the potential for advancement in traditional face-to-face learning environments, access to digital learning has been facilitated by an increased use of smartphones and tablets (Wang, 2016). The incorporation of these devices into the classroom has been found to support student engagement and independent learning (Rikala, 2013) with the added benefit of being potential sources of new data; everything from taps and swipes to data from applications implementing EDM methods can provide a wealth of knowledge on student performance.

In the late 2000s, the first company to develop adaptive, personalized learning technology intended for the K-12 classroom entered the market (Gunawardena, 2017). In addition to adaptive learning technologies that can be used in the classroom setting, online courses are now offered by a majority of school districts in the U.S. for two purposes: first, to offer expanded course selections to advanced students who have interests and abilities beyond what their physical school is able to provide; and second, to offer credit recovery courses to students who are at-risk or failing (Blazer, 2009). Research has shown that online courses have the potential to be as effective or even more effective than traditional instruction; however, due to the enormous variety of offerings and a similarly wide range of outcomes, success is heavily dependent on the quality of the program over the medium through which it is offered (Blazer, 2009). Because measuring actual learning is difficult in an environment that lacks specialist oversight and observation, Educational Data Mining and Learning Analytics methodologies can be applied to online classrooms to measure learning, identify trends, and plan interventions and improvements with greater accuracy than simply relying on grades and student feedback (Hung, Hsu, & Rice, 2012).

While data on traditional achievement measurements such as test scores and grades are important, non-cognitive, social, and emotional data are even better predictors of success in adult life (Cranston, 2016). The U.S. Department of Education Office of Educational Technology published a report focusing on characteristics such as grit, tenacity, and perseverance. In their 2013 report, they examine each of these characteristics and discuss how they can be measured and cultivated through the use of educational technology. The report collected data on mood through the use of a camera that measured smiles, sensors on student's chairs associated movement and slouching with boredom and fatigue, a pressure sensor on the student's mouse measured frustration through grip pressure, and a skin sensor measured levels of stress (Roberts-Mahoney et al., 2016). Through these biometric sensors, researchers were able to measure emotional states and associate them with non-cognitive characteristics in order to provide recommendations for fostering desirable student characteristics in schools (Shechtman et al., 2013). The report then continues to offer examples of curricula like school readiness and interventional programs that can be combined with their predictive model to help reframe mindsets and increase student resilience (Shechtman et al., 2013).

Another non-traditional data source has emerged with the proliferation of online communication methods like email, social media, and online chats and forums. Students, parents, teachers, schools, state agencies, and even the U.S. Department of Education are all interacting on social media in a manner that was previously unimaginable before the explosion of platforms such as Facebook and Twitter (Wang, 2016). Not only is social media a valuable platform for schools and agencies to maintain visible accountability and clear communication with stakeholders, but also the data from interactions on these platforms can be used to analyze public perception. Forums and chat rooms in course management systems such as Blackboard provide many of the same text data and are also valuable sources of data on social interactions (Wang, 2016).

Concerns

For the myriad exciting and positive implications of the use of big data in K-12 education, there are an equal number of serious considerations to be made regarding the potential for misuse. The most obvious of these is data privacy and security. In the private sector which has arguably been using big data more extensively than educational organizations, there have been an alarming number of data breaches in recent years (Mele, 2018), some even impacting children (Grant, 2018). In the school setting, educational data can sometimes be spread through multiple locations and tools, some of which may not interact well together; while this can make data more difficult for bad actors access, it can also make access much more difficult for those who should have it (Wang, 2016). While there is an obvious need to secure educational data from those outside the system, the potential for abuse or misuse of student data by educators and administrators is less obvious. For example, a 2014 study by Mayer-Schönberger and Cukier showed that data on student performance were "used not to push students to excel, but to push them out of higher education" (Wang, 2016).

The abuse of student performance data, whether stemming from intentional or unintentional bias, would undeniably result in an undermining of the goals of data programs in schools. The World Bank's Michael Trucano argues the presence of a "Matthew Effect" in education and technology – roughly translating to a theory that "the rich get richer while the poor get poorer." Schools in higher income areas tend to be plied not only with well-maintained devices and technology, but also with teachers and staff who are technologically literate. The students at these schools are also more likely to have access to these same devices at home, so they will already be familiar with much of the technology before classroom instruction even begins. While this level of access and literacy is certainly desirable for all schools, when these "richer" schools exist in the same system the "poorer" schools who have neither equal access nor equal skill, the result is widening of the achievement gap (Trucano, 2013). This theory highlights the need for proper planning before technological investments are made to prevent desired theoretical outcomes from being reversed altogether in practice.

Even when a one-to-one ratio of students to computing devices is available, the degree to which teachers incorporate them into their classroom activities is dependent on their own attitudes and beliefs (Penuel, 2006). Teachers who are primarily concerned that their students will use computers for playing games or surfing the internet have been seen to overlook the availability of these technologies in their lesson plans (Penuel, 2006). However, in studies where support such as assistance with lesson plans was provided, teachers invested more time into using technology and even reported that their expectations for student achievement had been exceeded and they were able to assign more challenging work on subsequent activities (Penuel, 2006). This study provides a clear illustration that teacher development and support programs are vital to the success of data-driven technology programs in schools.

One study from an early adopter of technology-rich data in schools described a major hurdle for their project in saying that "allocating the human resource required to collect, organize, and visualize student data was a real challenge for schools" (Mitchell & Conrad, 2003). As more volume and variety of data become available and technologies become increasingly complex, the task set before teachers to truly integrate data-driven decisions into all areas and subject becomes progressively more challenging (Hubbard, Datnow, & Pruyn, 2014). The necessity of providing data technology implementation training and support to educators is clear.

In both the public and private sector, demand for data scientists is high. Data-savvy businesses in the public sector employ not only data scientists who can extract insights from huge amounts of data, but also "translators who combine data savvy with industry and functional expertise" (McKinsey, 2016). Already, school districts have begun to outsource the collection, storage and analysis of education data (Roberts-Mahoney et al., 2016), but there may be additional needs. A 2014 study published by the Bill and Melinda Gates Foundation highlighted some teacher-identified challenges with tools that support data-driven instruction. The study found that teachers often struggled with inconsistent and untrustworthy data from the technologies used in their schools, and that they also found the large amount of data coming from disparate sources to be overwhelming and time-consuming. Ultimately, the study concludes with several recommendations to improve the ability to incorporate data-rich practices into classrooms; among them is to "invest[ing] in the staff, training, dedicated time, and professional development needed to integrate tools and practice" (Gates, 2014). If there is anything to be learned from the private sector, perhaps consideration of adding new roles to support teams which could be expanded to include data scientists, translators, and database managers is warranted.

While there is much to learn from successes in the private sector, not all private influence is positive. A 2016 publication by Roberts-Mahoney, Means, and Garrison examined twelve papers from government, corporate, and scholarly sources and found an alarming emphasis not on social or progressive goals, but on economic growth and corporate interest. The publication states that all papers they examined shared themes that "the purpose of education… is to augment human capital and train workers to develop twenty-first century skills." Roberts-Mahoney et al. suggest this reframing of the purpose of education through a corporate lens is due to considerable private lobbying, often that of conservative actors such as the Koch brothers, having a heavy influence on decision- and policy-making in American public schools. With increased control transferring to the private sector in policy and the influx of software and tools designed by corporations, special attention must be paid to ensure students are prepared not just to enter the workforce, but also to become successful members of a democratic society (Roberts-Mahoney et al., 2016). This corporate-guided shift in educational focus has also had the effect of deemphasizing non-cognitive skills. Even the 2013 report from the U.S. Department of Education Office of Educational Technology on qualities such as grit as indicators of student success reduced these qualities to data points that could be observed by sensors and cameras through interaction with a computer, not by teachers, parents, and staff through social interaction (Shechtman et al., 2013, Roberts-Mahoney et al., 2016).

As educational software and tools are integrated into traditional classrooms, the familiar boundaries around the role of the teacher can begin to lose focus. There is a real concern that adaptive learning technologies, while personalized, fast, and economical, "de-emphasize the professional knowledge and experience of teachers in relation to noneducators and to technology" (Roberts-Mahoney et al., 2016). When software designed by non-educators is doing the "teaching," the traditional role of teacher as instructor pivots to facilitator, data analyst, and guide (Roberts-Mahoney et al., 2016). This brings to mind two questions: first, to what extent is it a good idea to minimize the expertise and experience of teachers; and second, how well do we trust this technology?

In summary

Schools have been concerned with using data to improve student outcomes for decades, but never has so much data been available. The volume and variety of data continue to increase as new data sources are made available through devices such as tablets and smartphones, social media platforms, and educational course management systems. As new technologies are created, the potential applications for these data are endless. The study and analysis of big data is still an emerging field. The disciplines of Educational Data Mining and Learning Analytics help provide a framework for harnessing the power of big data and channeling it into insights. These insights can be used to create automated and adaptive courses and tests, or they can be shared with stakeholders to empower educators and students to make informed, intelligent decisions pertaining to their educational goals. However, great care must be taken that these data are used safely and always in service of improving the outcome of the whole student, not just as a future employee, but as an integrated member of democratic society.

References

Gunawardena, A. (2017). Brief Survey of Analytics in K12 and Higher Education. *International Journal on Innovations in Online Education*. DOI:10.1615/intJInnovOnlineEdu.v1.i1.80

Corbett, A.T. and Anderson, J.R. (1994). Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge. *User Modeling and User-Adapted Interaction, 4*(4), pp.253-278. https://doi-org.libproxy.utdallas.edu/10.1007/BF01099821

Mitchell, D., and Conrad, B. (2003). Using Technology-Rich Data Support Tools to Enhance School Improvement Initiatives in Chicago Public Schools: Are the Children Learning? How Do We Know?. Washington, D.C.: Distributed by ERIC Clearinghouse.

Foster, I., Ghani, R., Jarmin, R.S., Kreuter, F., & Lane, J. (2017). *Big data and social science : a practical guide to methods and tools*. Boca Raton: CRC Press.

Roberts-Mahoney, H., Means, A., & Garrison, M. (2016). Netflixing Human Capital Development: Personalized Learning Technology and the Corporatization of K-12 Education. *Journal of Education Policy, 31*(4), pp.405–420. https://doi-org.libproxy.utdallas.edu/10.1080/02680939.2015.1132774

Shechtman, N., DeBarger, A.H., Dornsife, C., Rosier, S., and Yarnall, L. (2013). *Promoting Grit, Tenacity, and Perseverance: Critical Factors for Success in the 21st Century*. Washington, D.C.: Center for Technology in Learning/SRI International.

Liñán, L.C., and Pérez, J.Á. (2015). Educational Data Mining and Learning Analytics: differences, similarities, and time evolution. *RUSC. Universities and Knowledge Society Journal, 12*(3), pp.98-112. DOI: http://dx.doi.org/10.7238/rusc.v12i3.2515

Baker, R.S., & Inventado, P.S. (2014). Chapter 4: Educational data mining and learning analytics. In J. A. Larusson and B. White (eds.), *Learning Analytics: From Research to Practice*, DOI 10.1007/978-1-4614-3305-7_4

Romero, C., and Ventura, S. (2013). Data Mining in Education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), pp.12–27. DOI: 10.1002/widm.1075

AlShammari, I., Aldhafiri, M., & Al-Shammari, Z. (n.d.). A Meta-Analysis of Educational Data Mining on Improvements in Learning Outcomes. *College Student Journal, 47*(2), pp.326–333.

Rikala, J., Vesisenaho, M., and Mylläri, J. (2013). Actual and Potential Pedagogical Use of Tablets in Schools. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments, 9*(2), pp.113–131. https://doi.org/10.17011/ht/urn.201312042736

Blazer, C. (2009). Virtual Schools. Literature Review. Washington, D.C.: Distributed by ERIC Clearinghouse.

Hung, J., Hsu, Y., & Rice, K. (2012). Integrating Data Mining in Program Evaluation of K-12 Online Education. *Educational Technology & Society*, *15*(3), 27–41.

Cranston, A. (2016). Leveling the Playing Field through Social-Emotional Learning: Cultivating Grit in Expanded Learning Programs. *Leadership, 46*(1), pp.16–21. Retrieved from http://search.proquest.com/docview/1819846107/

Wang, Y. (2016). Big Opportunities and Big Concerns of Big Data in Education. *TechTrends: Linking Research and Practice to Improve Learning, 60*(4), pp.381–384. https://doi-org.libproxy.utdallas.edu/10.1007/s11528-016-0072-1

Mele, C. (2018, August 1). Data Breaches Keep Happening. So Why Don't You Do Something?. *The New York Times*. Retrieved from https://www.nytimes.com

Grant, K. B. (2018, April 24). Identity theft isn't just an adult problem. Kids are victims, too. CNBC. Retrieved from https://www.cnbc.com

Trucano, M. (2013, June 7). The Matthew Effect in Educational Technology. [Blog post]. Retrieved from http://blogs.worldbank.org.libproxy.utdallas.edu/edutech/matthew-effect-educational-technology

Penuel, W. R. (2006). Implementation and Effects of One-to-One Computing Initiatives: A Research Synthesis. *International Society for Technology in Education, 38*(3), pp.329-348. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.461.7049&rep=rep1&type=pdf

Hubbard, L., Datnow, A., & Pruyn, L. (2014). Multiple initiatives, multiple challenges: The promise and pitfalls of implementing data. *Studies in Educational Evaluation*, *42*(C), pp.54–62. https://doi.org/10.1016/j.stueduc.2013.10.003

McKinsey Global Institute. (2016). *The Age of Analytics: Competing in a Data-Driven World.* McKinsey & Company.

Bill and Melinda Gates Foundation. (2014). *Teachers Know Best: Making Data Work for Teachers and Students*. Retrieved from http://k12education.gatesfoundation.org/download/?Num=2335&filename=TeachersKnowBest-MakingDataWork.compressed.pdf