Learning Analytics: Coming Out From Behind the Curtains

Philip Long | Chief Innovation Officer, University of Texas Austin

As Big Data and analytics become increasingly central to the management of both the academic and bureaucratic arms of postsecondary institutions, it's critical for leaders to deal with issues around algorithmic bias, transparency, data ownership and more.

Learning analytics has become a major focus of attention in the last several years.[1] It has the brought the expectations of Big Data and its associated statistical modeling from the cloud to the campus. While much of it has been

repackaged academic analytics aimed at elucidating the factors influencing four-year graduation rates and year-to-year retention, it has nevertheless provided insights into the development of interventions that have begun to move the needle on student success.[2,3]

A lot of this progress has been through harnessing the data that institutions have collected but often failed to usefully mine. These analyses are improving navigation of curricular pathways and identifying students at risk of failure in bottleneck classes that derails successful progression in majors. Simply identifying what courses are needed for seniors to take to complete their final year, *and making them available,* has had a remarkably positive impact on completion.[4] Sometimes it's the little things.

However, an emerging concern is the collection and storage of student data external to the learning institution. Are systems running "out there" protecting this PII (personally identifiable information) and using it for purposes "never intended for use by the school"?[5] Issues like this motivated academics and administrators from universities around the world to assemble on the California coast last year and develop a framework to inform decisions about appropriate use of data and technology in learning research for higher education.[6] This concern has also inspired the articulation of learning data and analytics principles by the Learning Analytics Community of Practice sponsored by IMSGLOBAL.[7]

We are wrestling with some difficult questions. When students interact with the digital environment provided by the institution, who owns the resulting data? These interactions often generate contextual information about the students' digital behavior. Metadata are created in response to the students' interactions in digital learning environments. For example, a timestamp may be associated with a mouse click, appending to that recorded keystroke information about the previous page the student was on. Are these data generated by the students' interactions the university's system?

It may sound like splitting hairs but ownership of the data carries with it not only responsibilities but also the right to decide how it is used. As Rayane Alamuddin and colleagues wrote in *Student Data in the Digital Era*, "In an environment with unclear ownership and governance, the most prominent risk is overreach—that someone will take action that crosses an ethical line."[8]

We have traditionally acted as if the university owns the data collected when students use their campuses' learning management systems. How does the student express "ownership" of their data? In the context of student-generated work in a lab, for example, student ownership of their submitted lab findings is their intellectual property.[9] But if their data contributes to the analytics used to assess their performance, or select among personalized learning pathways, what should the student know about this and what options exist regarding the use of their data?

There is a strong asymmetry in the power relationship between the student and their educational institution. Should students provide informed consent regarding the collection, use, and disposition of data they generate by their learning interactions? If the data collected, and the analytics applied to it, are to serve their learning and development and not just enable the institution to exercise its interests in maximizing a return on its investment in the learner, the learner should have some understanding of and say over their data.[10]

Another aspect of data ownership is understanding how it is used. What are the algorithms that are applied to student data, and what are the algorithms doing? The Learning Analytics and Data Key Principles declaration (Image 1) addresses this by asserting the importance of *transparency*. The goal here is to raise the visibility of the data trails students leave and the ways these trails are used, analyzed and impact them.

8. **Transparency**: Individuals have the right to understand the specific reasons, methods, and purposes for which their learning data is collected, used, and transformed. This includes any learning data being shared with third-party service providers and other institutional affiliates or partners. Individuals also have the right to know how their data is transformed and/or used thru processes such as summative or algorithmic modifications, particular outputs, and visualizations.

Image 1: IMS Global Learning Analytics & Key Data Principles

Is there reason for concern? Let's look at an example. In the introduction to Cathy O'Neal's wonderful book "Weapons of Math Destruction" she recounts the story of fifth grade teacher, Sarah Wysocki, working in Washington D.C. where only half of high schoolers were graduating and only 8 percent of eighth graders were reading at grade level.[11] A new Chancellor was hired to fix these problems and contracted a consulting firm to develop a teacher assessment tool called IMPACT to identify good versus bad teachers so they could get rid of the bad ones. Wysocki was off to a good start and had excellent reviews from the parents of her students and her supervisors. Yet her IMPACT score was in the bottom 10 percent and, as a result, she was let go.

Her IMPACT score was based on an algorithm that included many variables, none of which were shared or explained. Her students, a class of between of about 25, living in a poor southeast Washington district, likewise faced all manner of challenges besides coursework. Further the algorithm was subjected to little or no training data with which to tune its accuracy. Finally, it turned out that the end of year tests which formed the baseline against which her students' performance was determined turned out to have a lot of erasures, suggesting cheating to inflate the scores. When the district was confronted with this evidence they agreed it was suggestive of some issues but not conclusive. Wysocki's firing was upheld. O'Neal explained, "An algorithm processes a slew of statistics and comes up with a probability that a certain person might be a bad hire, a risky borrower, a terrorist, or a miserable teacher. That probability is distilled into a score, which can turn someone's life upside down. And yet when the person fights back, "suggestive" countervailing evidence simply won't cut it. The case must be ironclad."[12] The algorithm prevailed.

Other examples of "algorithmic bias" are emerging. Joy Buolamwini, a researcher at MIT's Media Lab, describes facial recognition systems that fail to consistently recognize people in particular ethnic groups. The code in these systems is drawn from reusable code libraries that developers frequently use to save time. But they exhibit what she calls "coded gaze," the bias embedded into coded systems and propagated by those who have the power to write the algorithms that go into them.[13]

These two examples, from distinct domains, reflect the power of algorithms and the lurking danger that can inadvertently, mistakenly or even intentionally emerge. The call for transparency in learning analytics is to raise our gaze and

ensure that biases do not encroach on an enterprise that so deeply influences the lives, success, and future of our students.

Much remains to be done. The conversation about the ethics, ownership and practices of learning analytics is young. It is likely that simply declaring that students should own their data fails to arm us with the insights and understanding to guide our actions. Asking for transparency in how we apply algorithms to students' data so students and their parents understand the purpose of these algorithms is a first step. But it, too, is not enough. We must broaden the view to include a wider context or put at risk a promising future.

_ _ _ _

References

[1] Phil Long & George Siemens (2011), "Penetrating the Fog, Analytics in Learning and Education", Educause Review, Sept-Oct, pgs:31-40.

[2] Philip J. Goldstein & Richard N. Katz (2005), "Academic Analytics: Uses of Management Information and Technology in Higher Education", ECAR, Res. Study 8, 113 pgs.

[3] UT Austin Leading National Efforts to Support Lower-income Students, UT News, last accessed Feb. 15, 2017: https://news.utexas.edu/2017/02/15/ut-austin-leading-efforts-to-support-lower-income-students

[4] Ohio State University, "OSU's College of Liberal Arts to offer four-year graduation guarantee to incoming students", News and Research Communications 02/21/2017, http://bit.ly/2lxwuFm

[5] Daniel Solove, "Interview with Kathleen Styles, Chief Privacy Officer, U.S. Department of Education" LinkedIn: http://bit.ly/2lxoILD , April 17, 2013

[6] Asilomar II: Student Data and Records in the Digital Era, Co-hosted by Stanford University and Ithaka S+R, last accessed Feb. 6, 2017: https://sites.stanford.edu/asilomar/

[7] IMSGLOBAL, IMS Global Learning Data & Analytics Key Principles, last accessed Feb. 6, 2017, https://www.imsglobal.org/learning-data-analytics-key-principles

[8] Alamuddin, R., Brown, J., & Kurzweil, M. (2016, September 6). *Student Data in the Digital Era: An Overview of Current Practices*. https://doi.org/10.18665/sr.283890

[9] Rule 90101: Intellectual Property, Section 6: Students and Intellectual Property. The University of Texas System, last accessed: Mar. 6, 2017 http://bit.ly/2lxCZs2

[10] Slade, S., and Prinsloo,P. (2013), Learning Analytics: Ethical Issues and Dilemmas, American Behavioral Scientist, 57(10) 1510–1529, DOI: 10.1177/0002764213479366

[11] O'Neil, Cathy (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy . New York, Crown/Archetype

[12] *Ibid*

[13] Joy Buolamwini, (2017), "InCoding — In The Beginning", last accessed Mar. 10, 2017,

https://medium.com/mit-media-lab/incoding-in-the-beginning-4e2a5c51a45d#.jlz2xt8wo

https://evolllution.com/technology/metrics/learning-analytics-coming-out-from-behind-the-curtains/