UC Irvine

UC Irvine Previously Published Works

Title

Contexts, critiques, and consequences: A discussion about educational data mining and learning analytics

Permalink

<https://escholarship.org/uc/item/9nw936fd>

Journal

Proceedings of the Association for Information Science and Technology, 55(1)

ISSN

2373-9231

Authors

Jones, Kyle ML
McCoy, Chase
Crooks, Roderic
et al.

Publication Date

2018

DOI

10.1002/pra2.2018.14505501085

Peer reviewed

Contexts, Critiques, and Consequences: A Discussion about Educational Data Mining and Learning Analytics

Kyle M. L. Jones

Indiana University-Indianapolis (IUPUI), United States, kmlj@iupui.edu

Chase McCoy

Indiana University-Bloomington, United States, chamccoy@indiana.edu

Roderic Crooks

University of California-Irvine, United States, crooksr@uci.edu

Amy VanScoy

University at Buffalo, United States, vanscoy@buffalo

ABSTRACT

The capture, aggregation, and analysis of student data is becoming ubiquitous at all levels of education—from primary to post-secondary—as institutions increase their adoption of information technologies to serve administrative and educational ends. In part, this has led to a growing interest among education scientists and administrators in strategically mining and analyzing troves of data in order to uncover student behaviors and intervene in student life. As is the case in other contexts where actors are pursuing Big Data’s supposed benefits, educational data mining is fraught with moral, ethical, and political conflict. The panel is composed of four researchers who analyze and critique educational data mining practices and learning analytics initiatives in particular micro- and macro-contexts, including academic libraries and professional advising, urban schools, graduate-level online learning, and higher education generally. Each panelist also represents a unique conceptual background, pulling from work in information ethics and policy, critical data studies, documentation studies, and higher education policy to empirically analyze and critically evaluate tools, systems, practices, policies, and values. Through their individual but thematically intertwined perspectives, the panelists will present their research and lead audience members in discussion.

KEYWORDS

Educational data mining, learning analytics, critical data studies, documentation studies, information policy

INTRODUCTION

Scholars, reporters, and government actors are increasingly playing close attention to the emergence of educational data mining (EDM) strategies and their application in learning analytics (LA) initiatives. EDM applies automated methods (e.g., machine learning) to learner data in order to surface, describe, and predict learning behaviors in particular systems (Baker & Siemens, 2013). LA takes EDM approaches a step further by applying these methods to investigate, understand, and optimize learning environments and

processes within an institutional context, not just a singular technological system (Siemens, 2012). Both EDM and LA require access to significant stores of data, some of which is focused on student learning while other data captures the minutiae of student life outside the classroom. EDM and LA have both benefited from the “data explosion” in education, which was sparked by the evolution of educational environments where information systems have become ubiquitous and various forms of data capture are increasingly common (Long & Siemens, 2011, p. 32). Due to these new technological conditions, student behaviors once “unseen, unnoticed, and therefore unactionable” (Bienkowski, Feng, & Means, 2012, p. ix) can now be described and analyzed, including a student’s social, educational, and increasingly mental state. New opportunities brought about by data aggregation and analytics has spurred some to claim that big data has arrived in educational institutions (Picciano, 2012). Claiming that it holds the power to transform “existing processes of administration, teaching, learning, [and] academic work” (Daniel, 2015, p. 913) in schools and universities, some see big data in education as a possible panacea for solving the seemingly “intractable challenges” (Campbell, DeBlois, & Oblinger, 2007, p. 42) education institutions face in an accountability culture that seeks to maximize outcomes with minimal cost expenditures.

When all these things are considered, it becomes clear that there a host of ethical and social issues with EDM and LA worth unpacking and critiquing. This is especially true given that the stoutest proponents argue that “there are no more excuses” (Crow, 2012) for *not* analyzing available data; others state that there exists a moral obligation to pursue data analytics given the availability of and access to datasets—to not use existing data would abdicate an institution’s responsibility to its students. If we accept this argument—that data *must* be collected, analyzed, acted upon—then we are led to ask some important questions: Who is collecting the data, how are they collecting it, and what right do they have to do so? Are the analytical methods valid and verifiable? To what ends are data analytics put, and are these ends justifiable? These questions and more surface important aspects of EDM and LA in need of targeted critiques of ideological, epistemological, and ontological perspectives aligned with big data in education. It is also important to take note of how EDM and LA become dominant methods and ideologies in a given context (e.g., primary education, higher education) and micro-context (e.g., charter schools, academic libraries), as the intersection of tools, norms, values, ethics, and goals will all vary.

This panel will examine educational data mining and learning analytics toward the end of raising awareness around related sociotechnical and ethical questions worthy of analysis using empirical, philosophical, and critical methods and approaches. The panelists will model these approaches by highlighting the unique but thematically intertwined work they have accomplished to date. The following sections provide an overview of each researcher’s approach before summarizing a particular project the researcher will discuss on the panel. The panel description ends with the structure of the panel.

INFORMATION FLOWS, PRACTICES, AND PROFESSIONAL ETHICS IN MICRO-CONTEXTS

In the context of higher education, administrators view the systemic application of data analytics as a means to support strategic retention, graduation, and time-to-degree improvement programs. To aid administrative ends and use these data and related analytics to better understand and report on the institution as a whole, it is increasingly the case that other institutional actors are being forced—or heavily “encouraged”—to begin strategically collecting student data and acting on learning analytics. Often, these actors are affiliated with micro-contexts, or groups of professionally affiliated actors, like advisors, librarians, and educational technologists. Each micro-context has its own norms, principles, and goals for which it creates and sustains supportive information flows (Nissenbaum, 2010). But the

degree to which one micro-context values and builds information flows to facilitate learning analytics may not align with another micro-context's expectations. Misalignment can create conflict in ways that enable one micro-context to suppress the professional ethics commitments and values that characterize another micro-context.

AFFECT AND POLITICS IN DATA VISUALIZATION

Despite the technical and conceptual heterogeneity of data accumulated in the various processes that constitute formal education, this portion of the presentation will point to the homogeneity of visual modes of interpretation used in unpacking and understanding data, particularly the construction and circulation of dashboards. Although data professionals themselves express a variety of positions with respect to the representational capacity of data, these important ambiguities are necessarily obscured by workflows that take visualizations not as (re)mediations or representations of data, but as objective, self-evident measurements. Sociology of quantification argues that data visualizations evoke feelings and emotions from their viewers, a dynamic at odds with the perceived objectivity and certainty claimed for educational data by researchers and policymakers alike (Kennedy & Hill, 2016). With this conceptual tension in mind, this portion of the presentation seeks to better understand the action of visualization via the method of semantic analysis (Kennedy et al., 2016). This method combines analysis of several typical data dashboards with interview data collected from these visualizations' authors, data professionals employed by a Charter Management Organization (CMO) that operates one dozen schools in South and East Los Angeles. This panel will discuss the initial findings in terms of political issues for students and communities so represented via dashboards and theoretical implications for the application of data science to the domain of public education.

DOCUMENTATION STUDIES AND CRITICAL LEARNING ANALYTICS

Critics of learning analytics have primarily focused on issues of student privacy and surveillance. Yet, there are other critical areas that need to be addressed in order to uncover the unintended consequences of learning analytics in practice. Although learning analytics advocates have often treated student data as unproblematic by promoting, often implicitly, data as objective and value neutral, this view is untenable. Modern documentation studies provide a conceptualize lens for critical assessing learning analytics practices. Framing data as documents and using the concepts of "made into" and "considered as" (Buckland, 2013) highlights data as socially constructed and subjective representations of students. Using documentation studies as a framework, four areas of future critical research are presented. These areas include data as not pre-factual objects; data patterns, apophenia, and data misjudgments; data categories and discriminatory institutional practices; and data doubles and student's right to access their data. Existing learning analytics cases will be examined using these critical areas to provide examples of how these areas can be applied in future critical learning analytics research.

STUDENT DATA PRIVACY IN HIGHER EDUCATION: THE ROLE OF THE SYLLABUS

As the ubiquity of learning analytics increases in higher education, it will be increasingly important for institutions to communicate how they are collecting data, to whom they are making it available, and to what ends the data are being used. Often, these communications are done with policy documents and notifications. But central to this discussion are instructors, who are in a unique position to communicate with and educate students about uses of their data. Such discussions may occur face-to-face, but it is often the case that instructors communicate about important issues—such as privacy—in syllabi. Conceptualizing the syllabus not only as communication tool, but also as an expression of

instructor values raises critical ethical and pedagogical questions about the role of instructors in students' understanding and value of their personal data.

PANELISTS AND TOPICS

Topic 1: Learning Analytics in the Academic Library

Kyle M. L. Jones is an assistant professor at Indiana University-Indianapolis (IUPUI) in the School of Informatics and Computing within the Department of Library and Information Science. His research centers on information ethics and policy issues associated with learning analytics tools, practices, and interests in higher education. Among other things, his past work considered how learning analytics violates student privacy rights and expectations. More recently, he has focused on the use of learning analytics in particular higher education micro-contexts. He asks questions related to how specific professional communities consider and reconcile data mining practices with extant values and norms, using case study and interview methods to surface data. Using this data, he examines effects from contextual nesting, conflict, and suppression to highlight how the application of learning analytics affects professional practices and identities. In this panel, Jones will highlight an interview-based study of academic library professionals who provided their perspectives on learning analytics, the compatibilities and conflicts with professional ethics commitments (e.g., ALA Code of Ethics), and political tensions with institutional administrators.

Topic 2: Affect and Visualization in Urban Education

Roderic Crooks is assistant professor of Informatics at the University of California, Irvine. He studies information and communication technology use in minoritized communities through field-based methods, often in the context of public schools. His current project explores the application of data analytics to the complex and politically charged realm of urban education. Urban schools in Southern California—high-poverty, public schools that serve primarily black and Latino students in the central zones of the city—face pressure to address persistent conditions of inequality by making use of novel forms of digital technology, to make communities that are perceived to be technologically bereft into the beneficiaries of progress associated with the tech sector. This project examines the work of data professionals working in this area to question the intellectual commitments of “data science” as well as the political implications of data-centric management of public institutions.

Topic 3: Documentation Studies and Critical Learning Analytics

Chase McCoy is a doctoral student at Indiana University-Bloomington. His work focuses on the practices of higher education decision makers when engaging with information visualization dashboards. He explores these practices and their consequences by examining how these actors are embedded within broader institutional socio-technical networks, how these networks shape data-informed decision-making practices, and the unintended consequences that emerge when these actors engage in practices that differ from what is expected by dashboard designers and stakeholders. In this panel, he will present work on a collaborative conceptual project that draws on modern documentation studies to outline emerging areas for critical learning analytics research and for the broader educational data science community.

Topic 4: Student Data Privacy in Higher Education Syllabi

Amy VanScoy is an assistant professor at the University at Buffalo in the Department of Library and Information Studies. Her work focuses on how professional thinking influences professional practices. Using qualitative methods, such as interpretative phenomenological analysis, she has explored how values, attitudes, beliefs, and practice theories influence professional practice, in the context of information services, library programming, and instruction. VanScoy will discuss her study of library and information science (LIS) syllabi. LIS courses, with their goal of preparing graduates to be savvy and ethical consumers of information, seem an ideal context for educating students about data privacy issues. However, the study found that few syllabi address student data privacy and, when addressed, syllabi reveal knowledge gaps, an overfocus on legal aspects of students' privacy, and a focus on specific techniques, rather than a broad conceptual understanding of the issues.

THE PANEL'S STRUCTURE

The aim of the panel is twofold. First, the panelists seek to provide audience members a broad overview of educational data mining (EDM) and learning analytics (LA). To this end, Kyle M. L. Jones will provide a brief talk on the state-of-the-art with regard to practices, technologies, and politics. Second, the panelists seek to provide unique conceptual lenses through which the audience members will be able to see how to enter into this area for their own research; here, each panelist will give a short talk on their topic as detailed above. After these aims are accomplished, the panelists will lead the audience in discussion regarding EDM and LA generally and more specific related issues, opportunities, and methods for developing a research agenda in this area and recommending paths forward with regard to practice, technological design of data-based artifacts (e.g., visualizations, algorithms), and policy.

ACKNOWLEDGMENTS

The panelists thank the Information and Learning Sciences (InfoLearn) Special Interest Group for their sponsorship. InfoLearn welcomes researchers from the information and learning sciences, and it aims to "advance theories of learning, information-seeking, and learning systems' design" (About InfoLearn, 2018). The SIG fuses the "scholarly disciplines of information science and the learning sciences stand to enrich one another theoretically, methodologically, and, empirically."

REFERENCES

- About InfoLearn. (2018). Special interest group: Information and learning sciences (InfoLearn). Retrieved from <https://www.asist.org/SIG/SIGILS/about/>
- Baker, R., & Siemens, G. (2014). Educational data mining and learning analytics. In R. K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 253-272). Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9781139519526.016
- Bienkowski, M., Feng, M., & Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief* (U.S. Department of Education No. ED-04-CO-0040). Retrieved from <http://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf>
- Buckland, M. (2013). Document theory: an introduction. In M. Willer, A. J. Gilliland, & M. Tomic (Eds.), *Records, Archives and Memory: Selected Papers from the Conference and School on Records, Archives and Memory Studies* (pp. 223-237), University of Zadar, Zadar.

Campbell, J. P., DeBlois, P. B., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. *EDUCAUSE Review*, 42(4), 40-57. Retrieved from <http://www.educause.edu/ero/article/academic-analytics-new-tool-new-era>

Crow, M. (2012). 'No more excuses': Michael M. Crow on analytics. *EDUCAUSE Review*, 47(4), 14-22. Retrieved from <https://er.educause.edu/~media/files/article-downloads/erm1241p.pdf>.

Daniel, B. (2014). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904-920. doi: 10.1111/bjet.12230

Kennedy, H., & Hill, R.L. (2017). The pleasure and pain of visualizing data in times of data power. *Television and New Media* 18 (8), 1- 14. <https://doi.org/10.1177/1527476416667823>

Kennedy, H., Hill, R.L., Aiello, G. & Allen, W. (2016). The work that visualisation conventions do. *Information, Communication & Society* 19(6), 715-735. <https://doi.org/10.1080/1369118X.2016.1153126>

Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5), 30-40. Retrieved from <https://er.educause.edu/~media/files/article-downloads/erm1151.pdf>

Nissenbaum, H. (2010). Privacy in context: Technology, policy, and the integrity of social life. Stanford, CA: Stanford Law Books.

Picciano, A. G. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks*, 16(3), 9-20. Retrieved from <http://onlinelearningconsortium.org/jaln/v16n3/evolution-big-data-and-learning-analytics-american-higher-education>

Siemens, G. (2012). Learning analytics: Envisioning a research discipline and a domain of practice. *Proceeding in LAK '12: The 2nd International Conference on Learning Analytics and Knowledge* (pp. 4-8). New York, NY: ACM. doi: 10.1145/2330601.2330605