Marcia Rapchak*

With the proliferation of Learning Management Systems (LMS) and other online learning technologies comes the abundance of data about students and instructors and their interactions with these systems. These data and their uses raise several concerns related to privacy, ethics, and equity. Students and faculty are often unaware of data that are being collected and do not have the opportunity to opt-out of the collection. Just as with any data that uses machine-learning to predict information about people, learning analytics are not neutral; they have the risk of reinforcing racial bias and other biases. However, learning analytics can be used to identify where students are not receiving support or to identify instructional areas that are not generating engagement. Academic libraries have been using learning analytics to assess the value of the library, raising ethical questions. Vocational awe and the myth of library (and technology) neutrality are barriers to equitable use of learning analytics. Academic librarians can lead their campuses in critical conversations around using learning analytics and the systems that generate such analytics.

INTRODUCTION

Imagine you’re an incoming first-year student, excited and nervous about classes starting soon, about to meet with your advisor to discuss your intended major and the courses you will take. While you went to a high school that was under-resourced and did not always receive the grades you aspired to, you are excited to have a clean slate at a new university and be the first person in your family to attend higher education. You tell her advisor about your dreams of being an aeronautical engineer, and thus you would like to enter as an aerospace engineering major. Your advisor runs your information through an advising program called Navigate, and an alert appears on the screen. The advisor reviews the alert, turns to you, and says, “Hmm, I wonder if we might start you in a major that you’ll be more successful in. Have you thought about Communications instead?”

While the specifics of this encounter are fictionalized, the reality of risk assessment advising programs are not. They are part of the complex and interconnected world of learning analytics. Universities are increasingly using learning analytics in ways that dictate student, faculty, and advisor action. Navigate, the software described in the opening vignette, uses a risk assessment program that...
prediction system that relies on several factors about students to determine if a particular major is low, medium, or high risk for the student. These prescriptive learning analytics can perpetuate systems of exclusion that impact minoritized students at a much higher rate than white, wealthy students.

It is not too late to take a more critical approach to learning analytics as colleges and universities are in varying places of large-scale adoption for the predictive use of them, though it is important to note that nearly all our institutions are using third-party vendors who collect and generate learning analytics. Academic librarians are poised to take leadership roles in the discussions on campuses about the adoption and use of learning analytics. Our profession is well-versed in issues of privacy, data preservation, and data ethics, and while we may require more professional development to be fully prepared to use learning analytics in ethical ways, we can use our expertise in data ethics and, perhaps more importantly, our passion around equity and inclusion as it relates to data, to ensure that these issues are considered during deliberations about what learning analytics the institution will use and when. Moreover, for those of us engaged in instruction, we can make personal decisions to create a more equitable classroom when considering our use or disuse of learning analytics. This paper will describe what learning analytics are, provide a brief overview of what libraries have done with learning analytics, address important considerations and concerns in their use, and then describe the initiative at the author’s institution that helped to bring forward some frank conversations about learning analytics.

**LEARNING ANALYTICS**

Learning analytics are the digital traces left behind when students, instructors, and advisors interact with digital learning environments. Learning analytics can be descriptive, diagnostic, predictive, or prescriptive. Descriptive learning analysis reveals what has occurred and nothing more. A descriptive learning analytic may be that 60% of students received an A on an assignment. Most of us with access to learning management systems or even a simple gradebook would be able to gather this information. This, of course, does not indicate why 60% of the students received an A, but many instructors do use descriptive analytics to determine possible actions. For example, if an instructor sees that 80% of the students answered an exam question incorrectly, they may decide to provide more instruction about the concept covered in that question, may reword the question in the future, and/or may drop the question from the exam grade entirely, but all these decisions would be based on what the instructor determined was best.

Diagnostic learning analytics tell us why something happened and may require several datasets, comparisons, and statistics to complete. For example, we may conduct an analysis of those 60% of students who received an A on an exam and find that 98% of them had clicked through all the content for the available modules in the learning management system. Correlation does not equal causation, but this seems to indicate that there may be a relationship between clicking through the modules and receiving a high grade. Considering the face validity of this, we may feel confident after such an analysis saying that going through all the content in the available modules contributes to receiving an A on the assignment.

Predictive learning analytics tell us that something is likely to happen in the future. For example, we may see that a student has not gone through the materials in the available modules in the learning management system, and so we may then predict that they will not receive an A on the assignment because our learning analytics tell us that this is often the case with other students. Now, this is a situation where learning analytics may be unnecessary to indicate that this is likely, but it illustrates what may seem like a benign use of predictive learning analytics. It may be useful to share this (anonymized and aggregated) data with students to encourage behavior that would be more likely to lead to higher grades. However, even in this relatively unconcerning example, if we saw that a student did not watch every minute of our lectures online and then went to grade their essay, we may be primed to grade them more harshly. When we factor in other variables in our analyses, especially unchangeable ones like demographics, predictive learning analytics can become especially worrying.

Prescriptive learning analytics tell us what should be done in a situation. Prescriptive learning analytics would say that a student must complete an additional module in a course based on their prior semester GPA, for example, to be successful in that course. The opening vignette was an illustration of prescriptive learning analytics at work, and these may be the most obvious as far as having potentially concerning issues, but all learning analytics should be used with care because of the privacy and other ethical implications in their use.
Creating A More Equitable Future Through Critical Approaches To Learning Analytics

ACADEMIC LIBRARIES AND LEARNING ANALYTICS

As numerous studies have argued, academic libraries have been under increased pressure to demonstrate their value to their institutions and their communities. ACRL’s own *Value of Academic Libraries Report* by Megan Oakleaf argues that libraries must bring forth evidence that shows how they contribute to the goals of the institution. All higher education institutions rely on students for at least some of their funding, and the student experience is key for the livelihood of a college or university. Learning analytics provide libraries with a way to link library use and positive outcomes for students, like retention and higher GPAs, which help to show that the library, while most likely not actively recruiting students into a major, contributes to student success.

Jones et al. provide a thorough overview of some of the ways that libraries have used learning analytics, including research that uses data from the ILS; students’ interactions with librarians through reference, instruction, and other public services; and also student data provided by the institution, like GPA and demographic data. They also point out that most of these studies appear to be done without a consideration for student consent and students’ ability to opt-out of data collection. Oakleaf argues, though, that we should be shifting our professional concerns about data privacy because the data that libraries have access to often does not always provide the correct or the precise data needed to demonstrate the library’s impact on students. She claims that gaining access to learning analytics at the institutional level, including data provided by the institutions about individuals students, and integrating the library into this institutional system, libraries will be able to develop more sophisticated and meaningful research about their impact.

Jones and Hinchliffe state that learning analytics, despite the acknowledged ethical issues in their use, particularly as it relates to student privacy, can indeed provide insights into how the library contributes to student learning. However, they make the case that more education is required for information professionals to be able to do so in a way that adequately addresses privacy issues that arise in the use of learning analytics. This has led to the creation of the Prioritizing Privacy continuing education program that is funded by the IMLS, led by Hinchliffe and Jones, which focuses on “learning analytics, privacy theory, privacy-by-design principles, and research ethics.”

CONCERNS ABOUT LEARNING ANALYTICS

Students, by-and-large, trust academic libraries with their data. While academic libraries unfortunately have a history of using third-party vendors who do not have the same dedication to patron privacy as we do in the profession (look at LexisNexis’s relationship with ICE), we have seen calls for increased emphasis on negotiating for patron privacy with these vendors, as well as finding alternatives to these vendors. In an analysis of studies on libraries and learning analytics, Briney found that many of the studies did not include robust data management plans to address fully issues around data retention, anonymization, ethical consent practices, and more. Clearly, the data handling practices of libraries who are using learning analytics need significant improvement to align with professional values of patron privacy.

However, concerns about learning analytics go far beyond data management issues, especially when we look at predictive and prescriptive analytics. Learning analytics that direct students to particular resources and steer them away from others damages intellectual freedom. Moreover, this can create a culture of surveillance and fear, which could particularly impact students of color, who are already over-surveilled. As Paris, Reynolds, and McGowan argue, “Research shows that such monitoring… chills engagement; has disproportionately negative outcomes for women, trans, and non-binary folks, and people of color; contributes to the entrenchment of unwarranted surveillance technology in education; and contributes little to student learning.”

For librarians that are instructors or take an instructional role, learning analytics can give us information about important aggregate information—how long on average students spent on a page, how well they performed on an online tutorial, or how many minutes they watched an instructional video—and this can help us, along with focus groups and surveys, consider what content is connecting with students and what is not, which is helpful information. It’s important to recognize what information such learning analytics are not capturing, though. As Wise, Sarmiento, and Boothe describe, in doing something they term a Blank Box Analysis, we can...
Marcia Rapchak

see that learning analytics miss much of the data about our students, like “external factors from their life outside the tool that shape how and why they engage,” revealing “the relevant educational constructs and values that we should take into account, even when there is no data to speak to it.”

Along with not capturing data, it may capture inaccurate and incomplete data; as Jones and Salo point out, a student opening a document does not mean they interacted with it, and their not opening could mean they accessed it elsewhere.

Another important consideration is that data are not neutral—the ways that data are generated and the assumptions about what is valued in creating learning analytics exist in systems impacted by economic, political, and social structures. The data captured in learning management systems reflect values embodied in higher education learning, values that have a history in maintaining white supremacy. In adopting systems that take these values as natural, objective, and true, we perpetuate structures that have historically excluded students from backgrounds that are not white and middle-to-upper class. For example, as described in Archer and Prinsloo, predictive learning analytics are not explanatory models—they describe what is possible but not why it is. They can be used to identify “risk factors” that are very problematic in their application, like identifying being a particular race as a “risk factor.” It is not the person’s race that is the “risk factor,” but systemic racism and other interlocking systems of oppression that lead to members of minoritized groups being less likely to be academically successful, however, a predictive model does not uncover these underlying causes, nor does it illuminate the appropriate intervention. In this way, learning analytics follow the path of other systems and algorithms that purport to solve issues of human bias and then perpetuate discrimination, now with a false claim that there is objective data to support such decisions, like with hiring algorithms and search algorithms.

Our reaction to the data is, of course, also not neutral. Already proposed was the situation where an instructor would grade a student more harshly based on information about what they completed in an online course. In predictive learning analytics, instructors and advisors are being told by the system to steer students to particular majors. As The Markup reports in regards to the use of Navigate in higher education, “The scores, which are one of the first things a professor or administrator may see when pulling up a list of students, can leave advisers with an immediate and potentially life-changing impression of students and their prospects within a given major.” Overworked advisors and professors are given a system that perhaps they do not entirely understand, that may use incomplete or problematic data, and then rely on these systems to make decisions and judgements about students. Students, too, if identified as being at-risk, may experience stereotype threat and may choose to pursue areas of study that they believe will not challenge them as much as others.

ADDRESSING CONCERNS

Jones et al. make a number of recommendations regarding learning analytics that include navigating institutional pressures to use learning analytics, even when they contradict our professional ethics. Conversations about concerns regarding the scope and impact of learning analytics must happen at our institutions, and librarians can be leaders in such conversations; we must have policies and strategies in place, not just at our libraries, but at our institutions, to make sure that we are considering the downstream impacts of how and when we use learning analytics. This need for librarian involvement in institutional conversations regarding learning analytics was the impetus for a project at the University of Pittsburgh on responsible use of learning analytics.

Each year at the University of Pittsburgh, the provost hosts a “Year of…” that focuses on a topic of interdisciplinary interest to the university community and provides funding for small grants to create programming and other initiatives to explore this topic. In academic year 2021-22, the Year of Data and Society, chaired by Dr. Nora Mattern, asked for proposals that explored the societal and human impact of data and its use. As an instructional designer and library educator, the author was interested in the ethical issues that arise in using learning analytics in higher education, especially as they related to the role they could play in systemic issues like racism and classism. With a team made of the Associate Vice Provost of Student Affairs and staff from Pitt’s Teaching Center, the author proposed a series of talks around the opportunities allowed through learning analytics, while also taking time to acknowledge the challenges and concerns surrounding their use. All the sessions in the series would be appropriate for an academic library to host or co-host and could be inspiration for other initiatives.
In the first session, “The Ethical Use of Learning Analytics to Improve Student Success,” George Rehry from Indiana University’s Center for Learning Analytics and Student Success and Lizette Muñoz Rojas from Pitt’s Teaching Center discussed the four quadrants of the ethical framework surrounding the use of learning analytics: privacy, power, purpose, and policy. Using this framework helps faculty examine the ways that their use of learning analytics aligns with appropriate use of data. The session emphasized that faculty should consider student expectations about data privacy, along with who has decision-making authority about what happens to the data and why (power), how the data will be used (purpose), and how the use of the data aligns with policies at institutional and federal levels before making decisions about whether and how they implement learning analytics.31

The second session was a panel discussion that included Tinukwa Boulder, Director of Innovative Technologies and Online Learning at Pitt; Amanda Brodish, Vice Provost for Data Analytics at Pitt; Marsha Lovett, Associate Vice Provost for Educational Innovation and Learning Analytics at Carnegie Mellon University; and Chandralekha Singh, Director of the Discipline-Based Science Education Research Center at Pitt. Panelists discussed the implications of learning analytics on students within their higher education settings. Some panelists mentioned using learning analytics to drive research around increasing feelings of belonging in students in a course, especially students who have been historically underrepresented in those courses. Other panelists described using learning analytics at a programmatic level to increase student retention and to examine what courses tend to be a bottleneck for student matriculation. Emphasized in this session was approaching learning analytics with an asset-based approach and ensuring that the focus is on improving the learning experience while being transparent and mindful of structural inequities that can be perpetuated in these systems.32

The third session was intended for students. Robert Ackerman, Senior Manager of Data Analytics for the Teaching Center at Pitt, Chad Burton, Executive Director, Analytics and Project and Portfolio Management at Pitt, and Cenna Crosby, a Pitt graduate student in the School of Education discussed with students what data is collected about them, how it is used, and other pertinent information including the privacy and governance of their data. Students were then given a chance to ask questions to further understand the implications of the use and collection of their data. A session like this could be held at a library where students are able to learn more about how the library collects and uses their data and what control students may or may not have over their data when it comes to learning analytics.33

In the fourth session, “How Surveillance Capitalism Ate Education for Lunch,” Roxana Marachi, Professor of Education at San José State University, described how decision-making around educational technology in universities can be concerning, to say the least, especially when using third-party vendors. She argued that these third parties are not held to the ethical standards that educational researchers in universities are held to, and this means that student data is not always protected and can be used in ways that create data harms without any consequence for the private sector. Based on her prior work, Dr. Marachi also shared how looking at the problematic use of student data in K-12 spheres and the lack of regulation in that area can give us a sense of what may come to higher education. She describes the use of capitalistic enterprises in edtech as the “theranos-ing of education,” where promises are made regarding what edtech will achieve without evidence of success and without consideration of the potential or actual harms.34

This project has positioned Pitt to be more aware of the concerns surrounding learning analytics from multiple viewpoints: as an institution, as faculty, and as students. By becoming more aware of some of the ethical issues, there have been a number of direct impacts: the Teaching Center is considering future sessions of their own on learning analytics; the Teaching Center’s Strategic Plan for Academic Year 2022-2023 commits to continuing conversations with Pitt IT about data privacy; the Provost’s Office will monitor issues of access, governance, and opt-out options in the decision-making around learning analytics; and the Teaching Center and the Office of the Provost will also keep what was learned through these sessions in mind when reviewing contracts with vendors. Through the project, we were able to help others learn about learning analytics: what they are, how they can be used, as well as concerns about their use and implementation. We hope this initiative—a proof of concept that reflects the calls for such discussions in the literature—encourages librarians to lead their institutions in examining their current policies on what data is collected and how it is used to benefit the students. We also need to consider the educational technologies we are using and what third parties are doing with student and instructor
data, with an emphasis on what data is collected and how it is being used in ways that impact the lives of minoritized students. While there are many ways that learning analytics may be implemented, we must consider what uses, if any, benefit students and which may cause unintended, harmful consequences.

Some questions we can ask of our institutions about data generated from educational technology include:
1. What are our policies and/or the third-party vendor policies around data security, data preservation, data anonymization, and, eventually, data destruction?
2. How can we hold third-party systems accountable and ensure they are maintaining the ethical standards we expect of ourselves? If we can't, can we turn to home-grown solutions instead? Alternatively, could we look toward the library consortial model or initiatives like SPARC for strategies to collectively pressure edtech companies?
3. How are we keeping students informed about what data are being collected and how data are being used? Do students have the option to opt out of data collection?
4. If we are using predictive and prescriptive learning analytics, what data are and are not included in the models? Is the creator of these models transparent about how these models work? Are race, ability, nationality, gender identity, and class or other factors that can act as a proxy for these (like zip codes, first-generation status, name, and high school attended, among others) being weighted heavily to predict student success, effectively maintaining systemic bias?
5. If we are using predictive and prescriptive learning analytics, are they being tested for bias in an intersectional way rather than through the lens of a single identity?
6. What values are being encoded in our learning management systems? Do these reflect our true values regarding not only what is meaningful learning, but our dedication to social justice and anti-racism as well?
7. Are we effectively and comprehensively training instructors and advisors about what learning analytics can and cannot tell us, and also training them to consider unintended harms in their use?

Some questions we can ask ourselves if we are using learning analytics in our own teaching include:
1. Are we using learning analytics to inform us about our own practices instead of using them to make assumptions about individual students?
2. Are we being transparent with our students about what learning analytics we will be monitoring and how we will be using it to inform our practice?
3. Are we always keeping the Blank Boxes in mind and considering what we want students to accomplish in our courses that likely cannot be captured by learning analytics?
4. Are we relying entirely on learning analytics and/or third-party systems for assessment in some cases (like for participation, for example), and is this approach reasonable and fair?

CONCLUSION

This is not to say—necessarily, at least—that learning analytics should not be used or that academic librarians are failing to maintain their professional values if they adopt them. This is to say, though, that the work of scholars like Ruja Benjamin, Joy Buolamwini, Safiya Noble, Sasha Costanza-Chock, Christine O’Neil, Virginia Eubanks, and others who have emphasized the ways in which automated data can perpetuate racist, ableist, heteronormative, cisnormative, and classist systems—along with the concerns about privacy, surveillance, data security, and violations of intellectual freedom that have been raised in the LIS literature—must be considered in our use of learning analytics. While academic librarians are not the only members of a campus community that can and should raise the issue of responsible use of learning analytics, they are well positioned to engage in cross-disciplinary conversations, to highlight examples from libraries’ negotiation of vendor contracts, and to point to successful multi-institution initiatives that have created alternative, open-access library systems.

We cannot allow vocational awe to obscure the impact of our use of learning analytics; our use is not inherently good or neutral. Academic libraries should proceed with caution in conducting research that uses learning analytics to establish their value, not only because of potential privacy violations, but also because of biases that can be baked into learning analytics; we must recognize that learning analytics often cannot tell us the why or, at...
the very least, not the qualitative why (not to mention the tenuous correlation-as-causation arguments of many of these studies)). For individual librarians, using learning analytics in their own teaching can highlight areas where they might improve, but it is important to note that learning analytics cannot represent completely what we want our students to accomplish in their learning experience. Moreover, our very use of them could introduce “data-supported” biases and surveil students, especially minoritized students, in such a way that curtails their intellectual freedom. When reflecting on what cannot and what should not be quantified, the potential impacts on our students and their learning experience must be centered.

NOTES

8. Oakleaf.
11. Jones and Hinchliffe.
27. Feathers, "Major Universities Are Using Race as a ‘High Impact Predictor’ of Student Success."
28. Archer and Prinsloo, "Speaking the Unspoken in Learning Analytics."
30. and Prinsloo.
33. Robert Ackerman, Chad Burton, and Cenna Crosby, "Learning Analytics: What Students Should Know," Responsible Use of Learning Analytics Series. (Panel at the University of Pittsburgh, Pittsburgh, PA, March 17, 2022).
34. Roxana Marachi, "How Surveillance Capitalism Ate Education for Lunch,” Responsible Use of Learning Analytics Series. (Panel at the University of Pittsburgh, Pittsburgh, PA, March 24, 2022).
35. Costanza-Chock.
37. Rodriguez, "Understanding Library Impacts on Student Learning.”