8-6-2018

Data, data everywhere: Implications and considerations

Matthew D. Pistilli
Iowa State University, pistilli@iastate.edu

Data, Data Everywhere: Implications and Considerations

Matthew D. Pistilli, Ph.D., Iowa State University

The amount of data created in higher education is staggering. Every click anyone makes within a learning management system is logged. Every ping of a wireless internet drop is noted somewhere. Student demographic information is provided at the time of admission, along with past academic history. Financial information, via the FAFSA, is collected for most domestic students on a yearly basis. Scholarship applications include minable data. Student involvement databases note which students are involved at specific levels of every club or organization on campus. Students use ID cards to swipe into recreation centers, residence halls, dining centers, or help labs, and notations are made. Every grade earned, along with corresponding credit hours and courses, is stored. Even on a campus of only a few thousand students, the daily total of information created is likely in the realm of terabytes, and that is just the data points created by students.

Digesting all of this data, plus the myriad other points generated and collected in higher education, can be at best difficult, and, at worse, mystifying. It is at this point that learning analytics appears to be a viable solution for many campuses. Learning analytics is ultimately a very simple term for a very complex, involved, and often misunderstood process that combines various student metrics and institutional data, resulting in deep insights being generated. Linking data points together to gain deeper insight into phenomena is important, but to what end, and at what costs?

As the other authors in this volume have shown, there are a great number of positive outcomes that can be associated with the application of analytics. Data can be used to identify areas of challenge for institutions and students. Improved learning can be demonstrated, and in some cases, increased retention rates for students previously identified as being at-risk of even going to college, much less staying enrolled.

This volume has also identified many of the challenges associated with the implementation and adoption of learning analytics in higher education today. First, and foremost, the fact that once disparate data sets are now intertwined allows for new views of student profiles to be created. These profiles often bring to light things that only students know about themselves, and, sometimes, can even highlight areas previously unknown even to them. Further, with the amount of information now put into one place, data security risks increase, as well as concerns about student privacy, right-to-know, and being able to identify students just from just a few data points. However, it is also important to note that, as Hora (2018) and Clow (2013) do, even with the advent of large compilations of data, large amounts of information remain inaccessible to predictive modeling software. As such, the *complete* profile we look to create in an effort to identify and intervene with students before problems become real is never really, fully complete; it is only as complete as the data we input into it.

Other authors in this volume importantly discuss the notion of working with students and other involved stakeholders when constructing learning analytic environments, rather than trying to work for them. McKay's (2016) point that "we have to take advantage of the information age in which we find ourselves and dissolve the walls between research and practice in education" (p. 7) clearly and neatly underscores the fact that learning analytics cannot simply be a field of research, but, rather, one of practicality and pragmatism; one that allows researchers and students to meaningfully interact with one another to achieve the same end of student success. In the same vein, the interactions of researcher and student, or, more precisely at present, researcher and student *data*, also need to be addressed. Prinsloo and Slade demonstrate the challenges associated with ethical use of data – in other words, what should be done rather than what could be done with information – especially around the concept of students consenting in some way, shape, or form to allowing their data to be used, even if it is to potentially benefit them and their learning environments. In the previous chapter of this book, Klein and Hess take an in-depth look at some of the policy implications of LA and the associated challenges of using these data in assessment practice.

In the end, the ethics behind the use of data, never mind the actual collection of the data in the first place, cannot be ignored. There is great potential for good within these data. Different datasets can be merged, and, using learning analytics, analyzed in an effort to identify students who are underperforming or who might be at risk of not persisting into a second term or year of study. Similarly, a dataset could also indicate that specific courses or curricula create challenges for certain groups of students, regardless of instructor or the preparedness of those enrolled. The same data set could be used to identify instructors who struggle teaching different groups of students with the intention of developing skills and knowledge about pedagogy and curriculum to enhance their instructional methods. All three of these scenarios are in line with what Knight et al. (2018) discuss earlier in this volume: that implementing any form of analytics should not be done *to* users, but, rather, *with them* as partners in the process.

At the same time, there is great potential for harm in these data as well. Zook, et al. (2017) make this point bluntly: “data are people and can do harm” (p. 2). They continue, noting that “harm can … result when seemingly innocuous datasets about population-wide effects are used to shape the lives of individuals or stigmatize groups," especially when there is no means of recourse for people in those groups whose lives have been reshaped or stigmatized (p. 2). The self-fulfilling prophesy, wherein a student is identified as being at risk of doing poorly in college is told as such and then ultimately drops out of school, is always an ever-present concern as well. Additionally, not providing sufficient context with regard to how a risk marker is identified can have similar effects. As noted in Hora's chapter in this book, assuming that the collection of data will, in and of itself, create changes in pedagogy and practice is a fallacy. As will be discussed later in this chapter, institutions could potentially use LA to identify faculty or instructors whose courses or methods create challenges for student success.

How do campus administrators and faculty mitigate the power that data have in identifying areas for improvement and the need for students and instructors to be autonomous learning agents within our institutions? At what point does the use of data reach the point of sublimity, wherein student behavior is so prescribed that learning becomes rote and wholly lacking in critical thinking and application? These questions cannot be answered wholesale; each is highly nuanced, and the contexts in which learning analytics are applied will dictate the extent to which scenarios associated with these queries become reality or not. The point is that unless very difficult questions like these are grappled with by faculty, administrators, and students, the use of learning analytics on any campus can quickly become suspect.

To address this, and to conclude this volume, four main questions are considered:

- What does the advent and growth of learning analytics mean for students, faculty, and administrators?

- What issues will arise as more and more data is collected and pressed into service by an institution? What concerns will present themselves, and how might they be addressed?
- How can institutions use learning analytics in a way that enhances both the learning and the teaching process positively, and does not punish faculty or students for needing to improve?
- Ultimately, how should – not can – data be used, and to what ends?

Some of these questions will be answered fully in the remaining pages of this chapter; others will be posed with a framework for others to explore as they embark on their own paths of data use. Each will result in more questions than answers, requiring input from many facets of an institution to answer – assuming the questions can be answered at all. In the end, this chapter will provide insight into things that should be thought about before investing in or employing learning analytics.

What does the advent and growth of learning analytics mean for our students, faculty, and administrators?

The landscape of learning analytics tools is vast and growing rapidly, and many institutions are choosing to invest in some forms of technology that, in and of themselves, are analytics tools, or, at the very least, have some form of learning analytics capabilities built into them. But to what end? Institutions often note that the greatest reason for investing in analytics is to improve student retention (Arroway, Morgan, O'Keefe, & Yanosky, 2016). Arroway et al. (2016) continue, noting that the next four reasons for investment in learning analytics after student retention are improving students' course-level performance, demonstrating higher education's effectiveness, reducing time to degree, and understanding the characteristics of the students at their institutions. Of these five reasons, only one – improving course-level performance – is directly related to actual *learning*, something Gašević, Dawson, and Siemens (2015) remind us is the primary purpose of *learning analytics*.

Of the other four, one is trying to increase an outcome (student retention) without addressing its antecedent (student success), and another looks to reduce time to degree – something that intuitively is not always correlated with increased learning. Of the remaining two, student characteristics do not require advanced analytics to examine, and the ways in which we currently demonstrate higher education's effectiveness does not remotely fall in the realm of what learning analytics is designed to do: measure, collect, analyze, and report on data about learners, including where and what they are learning, in an effort to understand and optimize students' learning environments (Society of Learning Analytics Research, 2012).

The lack of focus on learners and what they are learning when measuring the effectiveness of higher education was recognized by the Boyer Commission (1998) when the authors noted that "most students graduate having accumulated whatever number of courses is required, but still lacking a coherent body of knowledge or any inkling as to how one sort of information might relate to others" (p.6). The authors continue, explaining how higher education institutions in general, and research universities in particular, were evaluated for effectiveness:

> The standing of a university is measured by the research productivity of its faculty; the place of a department within the university is determined by whether

> its members garner more or fewer research dollars and publish more or less noteworthy research than other departments; the stature of the individual within the department is judged by the quantity and quality of the scholarship produced. Every research university can point with pride to the able teachers within its ranks, but it is in research grants, books, articles, papers, and citations that every university defines its true worth. When students are considered, it is the graduate students that really matter; they are essential as research assistants on faculty projects, and their placement as post-doctoral fellows and new faculty reinforces the standing of the faculty that trained them (Boyer Commission, 1998, p. 7).

Almost twenty years later, the Spellings Commission (U. S. Department of Education, 2006) found that there "is a lack of clear, reliable information about the cost and quality of postsecondary institutions, along with a remarkable absence of accountability mechanisms to ensure that colleges succeed in educating students" (p. x). The authors continue, highlighting the great many "shortcomings of postsecondary institutions in everything from graduation rates and time to degree to learning outcomes and even core literacy skills" (p. 3). In short, the reasons for implementing learning analytics, at least as discussed in Arroway, et al. (2016), are limited and incomplete.

But what are *good* reasons for implementing analytics? This is the question that needs to be answered in order to address what the growth of learning analytics means for students, administrators, and faculty. Ultimately, the *good* reasons are those that allow the technology developed or purchased to do *what it was designed to do* and allows the institution to accomplish what it was wanting to – meaning a plan needed to be in place before investment in a tool. These two pieces do not always coincide.

All too often, we tend to try to make technology do things it was never designed to do, and then we get upset at it when it will not do what we want because what we want is not in its scope of abilities. Prior to investment in a technology, institutional decision makers need to meet with key stakeholders across campus – including faculty, administrators, student support staff, and students themselves – to determine what it is they want to try to accomplish through the combination of various sets of data. As I wrote with colleagues, successful implementations of learning analytics require forethought and reflection on the desired ends an institution is trying to be achieve through newly acquired or developed technology (Oster, Lonn, Pistilli, & Brown, 2016).

The other side of the learning analytics coin involves taking action; something has to be done with what was learned as a result of implementing analytics. Identifying students at risk of doing poorly in a class or dropping out of an institution with a certain degree of accuracy is important, but, absent an appropriate intervention, is insufficient. My colleagues and I termed this concept the *obligation of knowing* (Willis, Campbell, & Pistilli, 2013). In short, once something is known, what are the involved actors obligated to do as a result? These actors clearly involve the person or persons utilizing the learning analytics technology, but also include the students, instructors, and staff, themselves – and may also indicate actions that need to be taken by the university.

Overlaying all of this is the notion that, in general, faculty and staff concerned with academic success are working to identify some level of risk present in one or more groups of students with the intent of intervening so that those students can adjust their behaviors sufficiently to be successful. Just as knowing how information will be used once it is known is important to implementing learning analytics, so, too, is the use of language that will be used to describe and provided to students, and will be published about the implementation. Decisions must be made as to who will have access to the algorithms used and how they'll be explained to stakeholders. Further, what students or faculty will receive as a result of a learning analytics algorithm being implemented is also important. Will an index score be generated, and then provided to students or faculty? What context will accompany that score? How will it be interpreted for whomever is delivering that information or for the recipients themselves if there is no intermediary? Perhaps of most importance is the nomenclature that is assigned to students and is then used, particularly among advisors, administrators, and faculty.

All too often the phrase *at-risk student* is used. This is a problematic term, in that it defines the totality of a student's existence. Placing the term *at risk* before the student indicates that in all things, all activities, all behaviors, that student is at risk. At risk of what? It doesn't matter, since the term describes the whole student. Johnson (2017) describes this as a near "dehumanization [of students… wherein] institutional decisions are made based not on humanistic complexes of individual and social meaning but on mechanical processes of measurement, classification and response" (p. 79). Instead, the term *students at risk of [something]* should be used, where the something is what the learning analytics technology was created to identify as an area of concern for students. This could be performance in a given course, not getting involved or being socially isolated, leaving a major, or outright dropping out of an institution.

While the threat of reducing a student to a mere number or category is concerning, defining the totality of a group of students with a simple phrase creates a stereotype about those students and, as a result, introduces the predicament of stereotype threat. Steele and Aronson (1995) describe stereotype threat as "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (p. 797). In other words, in defining a broad group of students as being at risk, and then making that label known to instructors, advisors, and, possibly, the students themselves, this label will forever color the light in which they will always be viewed. This can lead not only to differential treatment or advisement, but also to students "suffering from self-fulfilling prophecies" (Pistilli, Willis, & Campbell, 2014, p. 91); that is, if they believe that they are likely to do poorly in a course, then they are likely to do poorly – especially if the stereotype is at all reinforced.

Willis et al. (2013) furthered this concept, noting that there is no real way of knowing how students (or faculty, or staff, for that matter) will react to the release of data. We posited that data "constructed in an … ethically-sound model" might actually result in a realm of empowerment for the student to take action in a way that creates favorable and positive outcomes (Pistilli et al., 2014, p. 91). Zook et al. (2017) echo this and take one step further, suggesting that organizations develop, in collaboration with researchers and those affected by these algorithms (e.g., students, faculty), their own codes of ethical conduct to follow and adhere to when building and applying these kinds of algorithms and models. However, for these kinds of outcomes to be achieved, careful consideration needs to be taken when developing or applying algorithms and ethical models, and, more importantly, context and courses of action need to be provided alongside whatever risk levels are identified for various groups of students.

What the advent and growth of learning analytics means for higher education, ultimately, is that data never before available or visible to the general user will now become prolific. Campuses need to work to manage the use of data so as to allow for the greatest amount of beneficence possible. Intentionality needs to be placed behind the purchase and use of the technology, the interventions employed, the language used in messaging, and the metrics used to determine the usefulness and success of the implementations. New levels of transparency surrounding data privacy and use also need to be present, and many campuses will have to thoroughly review their policies and practices surrounding the use of data to support student learning. While federal laws, including the Family Educational Rights and Privacy Act (FERPA; 1974), allow for the use of students' data to be used to supplement their own learning (provided that those receiving the data beyond the student have legitimate educational interests, which faculty, advisors, and student support staff generally do), many institutional policies, procedures, and legal counsels may interpret these laws more narrowly or may be in conflict with other policies and practices in place on campus.

In short, this question requires institutional actors to consider the following:

- the desired outcome(s) of a learning analytics implementation;
- granting access to the output of the system, and the formats that will be provided;
- nomenclature and language that will be consistently used around the use of learning analytics;
- interventions that will be employed, and staff/offices that will be affected by those interventions;
- the policies and practices surrounding privacy, data access, and the use of student data; and,

- assessment of effectiveness of the learning analytics system and how that information will be incorporated into future uses of the technology.

What issues will arise as more and more data is collected and pressed into service by an institution? What concerns will present themselves, and how might they be addressed?

At the core of learning analytics is the datasets that are mined to develop insights into students', instructors', or institutions' behaviors. As more and more data is generated by students or as a result of students interacting with a university, these "digital breadcrumbs" (Norris, 2011, p. 1) are becoming more and more prevalent. This brings the issue of data protection and privacy forward.

As discussed in the previous section, FERPA allows data to be used by those with a legitimate educational interest, and most institutions argue that a legitimate educational interest involves helping those paying tuition and fees to be successful in their educational pursuits. However, the policies and practices on most campuses are often in conflict with one another, particularly surrounding the notion of what data can be used by whom and for what purposes. Further, as more and more third-party vendors come forward with analytics, data combination and aggregation, and data storage solutions, very robust datasets – which oftentimes will include not just data about a student, but also about the instructors they had and the offices with which they interacted – will be sent to more and more places with varying levels of anonymity.

There are a few concerns here that individual institutions and the higher education community as a whole must wrestle with as student success takes hold as the driving forces behind these kinds of investment.

1. Not all student data or outcomes are equal.

   Different pieces of information about different students can be more predictive than others when trying to identify a potential outcome. For example, high school grade point averages are strong predictors of performance in college-level English classes, but standardized tests like the ACT predict math performance better than high school GPA, and the predictive power of GPA and standardized test scores varied by students' ethnic background, the time elapsed between graduating from high school and enrolling in college, and the urbanicity of students' hometowns (Bracco, et al., 2014; Hodara & Lewis, 2017). With that as the background, then, there are inherent challenges associated with using all forms of data. What are the limits on using proxies for low-income students with regard to identifying students at risk of doing poorly, especially when research largely bears this out? To what extent should first-generation status be used, particularly when there are varying and competing definitions about what first-generation actually means? Furthermore, what is the *correct* dependent variable to predict, and can the same algorithm accurately predict this for *all* kinds of students?

   Beyond these questions, there also exists the issue of how long various pieces of data are viable predictors of future behavior. To wit, are standardized tests and high school GPAs relevant after a certain amount of time? DeBerard, Spielmans, and Julka (2004) found that high school GPA was a very strong predictor of students' achievement in the second semester of college, and Stumpf and Stanley (2002) demonstrated a relationship between high school GPA, SAT scores, and eventual graduation from college. Stumpf and Stanley, however, did not examine any student achievement data once the students entered college, only that they graduated.

But at what point do some data become more important and other data less important, and do the algorithms employed take these data limitations into account? For example, Campbell and Oblinger (2007) posit that while high school GPA and pre-collegiate test scores may be strong predictors of how a student will perform academically during the first year of college, using real-time data may generate far more accurate models of student success. Adelman (2006) demonstrated this in *The Toolbox Revisited,* where he wrote that:

- students completing less than 20 credits in their first year of study were far less likely to graduate from college;
- students who enrolled in college for a summer term tended to have higher graduation rates; and,
- GPAs collected after the second year of college coursework strongly predicted persistence to degree

All of this is relevant data that supersedes the predictive power of high school achievement. However, none of this data is available prior to the events actually occurring. It is imperative that algorithms change and grow as the students, faculty, staff, and institutions change and grow. Furthermore, it may be the case that *one algorithm per campus is simply not enough*; where each incoming cohort of students can be different from the ones that preceded it, so too are the variables that predict or influence the dependent variable likely to change over time.

One-size-fits-all algorithms and tools cannot be used in the increasingly diverse and complex environments that are higher education institutions. Always looking for the same outcome using a static set of variables will yield results that, while actionable, aren't always meaningful for various groups of students. Institutions need to consider these limitations very carefully and deliberately. Further, they must decide how much misidentification of students at risk of a specific behavior or condition they are willing to tolerate. Failure to do so will likely result in students and other stakeholders demanding changes or no longer holding any stock in the outcomes, and likely deeming the system to be less than useful. If users perceive a system not to be useful, they will not use it (Davis, 1993).

2. Large sets of data can reveal information about things that were not originally intended to be found.

   Bienkowski, Feng, and Means (2012) note specifically that examining large sets of longitudinal data can "result in disclosure [that] may be hard to foresee" (p. 42). To wit, when hundreds of thousands of rows, each with several hundred data points, from a student information system (SIS) are combined with an equally dense set of data from a learning management system (LMS), the use of data mining techniques in conjunction with learning analytics implementation will reveal findings that were never part of the original reasons for implementing learning analytics. For example, while the intent may have been to identify students whose past academic history might require tutoring or other academic interventions to help them be successful in a given course, other information or unexpected outcomes might also emerge.

Imagine a student taking a full load of courses that include a math course and an English course. In the student's math course, the instructor is utilizing an analytics system the campus invested in to provide students with feedback as to how they are doing in the course and what they might do to maintain or improve their grades. One of the pieces of feedback provided to the student was to visit the math help lab, a space where graduate students and lecturers are available to explain concepts to students who are not fully grasping them by attending class or doing the reading. This student takes advantage of the math help lab and realizes an increase in homework and quiz scores as a result of better understanding the material. This same student also is struggling in her English course, and realizes that if the university is offering a math help lab, perhaps there also is an English help room she could visit. In searching for that help, the student also realizes a host of other academic resources available to students she had no inkling existed that she starts taking advantage of to improve her performance in all her coursework.

The help-seeking behavior described in the previous paragraph was driven by the application of analytics in one course. Analytics also has the power to identify patterns of use of various resources by students and the relationship between those patterns and performance in various courses or overall engagement with the campus community.

Presume another student in the same math course as the student in the previous paragraph. His intervention suggested that he speak with an academic advisor about his current schedule and overall performance in college. As he met with his advisor, several non-academic concerns were expressed by the student – his father had recently lost his job, he was having relationship troubles with his girlfriend, and he was struggling with his own mental health – all of which were contributing to his lower-than-average performance in math (where analytics were applied) as well in other courses. Given this knowledge, the advisor could refer the student appropriately – to financial aid to see if other aid were available, and to the counseling center to help address his relationship and mental health concerns. Thus, analytics can (and should) serve as a spotlight – highlighting obvious areas of concern (poor performance in a given course) that leads to an intervention (meeting with an academic advisor) resulting in a student being referred to appropriate resources (financial aid, counseling services). The hope here, as with the first student, is that by increasing help-seeking behavior for both academic and non-academic concerns, we are able to broadly increase overall academic achievement and improve students' ability to learn.

There are positive aspects of collegiate involvement that can be identified as well – or at least interpreted and applied in ways that can result in more positive outcomes. Perhaps by looking at when and/or where students are logging into the campus' LMS, their use of campus gymnasia or recreation facilities, and their course performance, interventions can be used to help students understand when the best times to do certain activities can be while still maintaining high academic achievement. Regardless of how positive or negative the unanticipated results are, institutions must have a plan in place for dealing with this information (obligation of knowing) once it has come to light.

In the end, however, Knight, Brozina, Kinoshita, Novoselich, Young & Grohs (2018) caution that student data should not simply be mined because it exists. Students indicated that they had very concrete ideas about which data points could and could not be accessed by various individuals in the academic enterprise. As I also propose later in this chapter, Knight et al. (2018) indicate that students should be made part of the decision making process behind which data are used, the extent to which those data are used, and who has access to both the individual data points *and* the aggregate outcomes of the analytics tool. Even if the intent of the analytics effort is to find previously unknown intersections of behaviors and outcomes, the desire to unearth these findings must be tempered with privacy and potential use beyond initial intent in mind.

3. Transparency is key to the success of the effort.

Multiple levels of transparency must be taken into account for efforts involving the use of data. First, individual students must be informed that their data will be used to help them, and other students, be successful. This is usually part of the FERPA or other data privacy disclosures that are provided via email to students at the beginning of academic years that, like other online privacy policies and terms of service, usually go unread (Milne & Culnan, 2004; Ober & Oeldorf-Hirsch, 2016; Waldman, 2017). Efforts should be made to directly inform students what data will be used and how it will be used in an attempt to help them be more academically successful. Prinsloo and Slade, in this book, make a strong case for working with those whose information will be used in analytics algorithms to determine how that data will be used and who will have access to it. Further, they conclude their chapter noting that "there remains a need for institutions to consider how to better engage and involve students in the provision of meaningful consent." (NB: The Open University in the United Kingdom (Online Student Support Services, 2016; see also http://www.open.ac.uk/students/charter/essential-documents/a-to-z#n340) has done some excellent work in this area, and other institutions would be well to follow suit.)

Second, faculty, administrators, and support staff need to be made aware that learning analytics are being employed, the rationale and intent behind the implementation, how the algorithms work, and the things they should individually and collectively do with regard to the output of the systems used. This aligns with the concept of readiness for learning analytics (Arnold, Lonn, & Pistilli 2014; Oster et al., 2016), where my colleagues and I posit that a successful implementation of analytics "requires the appropriate management and application of resources and personnel that results in the ability to effectively educate students and enhance the educational experience" (Oster et al., 2016, p. 172). If stakeholders are not ready for the advent of learning analytics (or any technology or system, for that matter) on a campus, that is, fully aware of what is being implemented and its potential usefulness, the reasons behind the implementation, and their subsequent role once the system is up and running, then the application of learning analytics will likely fail.

Helping key stakeholders understand how and why data is being used and what they should do with it can go a long way toward the acceptance of learning analytics on a campus. Ideally, stakeholders have been involved since the inception of the plans to use learning analytics at a given campus. Regardless of when they were made aware of the project, however, stakeholders need to be well-informed actors representing a wide swath of the institutions' faculty, staff, and students (Arnold et al., 2014). Per Johri's chapter, users also need to see the relevance of these data to their routines and practices.

Third, and as alluded to in the previous point, the algorithms employed should be as open as possible so as to demonstrate how and why a student was identified as being at risk of the predicted behavior or outcome. Most third-party LA products have a black-box model for their algorithms and output. This means that institutions upload a great deal of data into a system, which are crunched by the vendor and returned to the campus—a process that results in students having been identified as being at some level of risk. However, this is done without notifying campuses as to the parts of each student's record that caused a risk classification to be placed on students' records.

While the proprietary nature of these algorithms is understandable, if, as my colleagues and I, and others, have posited, that the point of using learning analytics is to create both "a shared vision [in] support of student success" (Norris, Baer, & Offerman, 2009; Oster et al., 2016, p. 264) and behavior change in students that meet that end, then not knowing what data points raised the flag of concern for the algorithm does not help an institution or its actors guide students toward making necessary alterations to their actions. Consider the following background information, and then three different scenarios where an algorithm might assign a risk level for the student as he enrolls in a specific course.

Background Information:

> *An African American male chemistry major with an overall GPA of 2.73 is looking to enroll in organic chemistry. The student is from out-of-state, is first-generation and Pell grant eligible, had a high school GPA of 2.4, and earned a composite ACT score of 23. Among other courses, he has taken the two chemistry courses and two math courses required as prerequisites for organic chemistry, earning a C in Chemistry I, a C+ in Chemistry II, a C in Calculus I, and a B in Calculus II, for a GPA of 2.325 in the prerequisite courses.*

Scenario 1

> *Suppose an algorithm places a heavy emphasis on past academic achievement in determining a risk level associated with students' potential success in organic chemistry. There is no way for students to mitigate how they performed in earlier chemistry courses or associated math courses, meaning that even if a student had consistently improved performance over time, as our example student did, the algorithm would see the relatively low prerequisite*

GPA earned by the student and would likely then assign him to a higher risk category, indicating that he would be likely to struggle in the organic chemistry class. Without knowing that the GPA was the reason for the assignment of the higher risk category, an advisor – who may have noticed the upward trend in grades and thought absent the algorithm's output that the student was ready for the next course – might work to dissuade the student from even enrolling in organic chemistry, thereby potentially altering the student's choice of major, and, possibly, continued enrollment at that institution.

Scenario 2

Alternatively, assume our student has a risk level assigned not because of past academic history, but, instead, for immutable demographic characteristics such as sex, race/ethnicity, socioeconomic status, or first-generation status. There is nothing that he can do to change those factors, but because other students with similar characteristics struggled in the same course in the past, this student is seen as being at risk of doing poorly.

Scenario 3

A third algorithm might indicate that our student is likely to do poorly because of a combination of lower grades in the past and failure to make use of university resources such as tutoring, help rooms, or Supplemental Instruction. Absent knowing that the lack

> *of use of these supports is the reason for the risk categorization, there is no way for the student to do anything to change his behavior to work towards a level of success higher than what the algorithm predicted.*

It would be next to impossible for an advisor or faculty member to correctly guide students toward behaviors that would potentially enhance their success given any of these three scenarios. How a student is identified as being at some level of risk for a given outcome must be evident to those who will be interpreting and providing that information back to students.

Fourth, transparency of the algorithms and data used to drive them should allow for implicit bias to be recognized and ameliorated. Thiel and Zimmaro (2017) note that data "*are not neutral*; [the assumptions made in data models] reflect the biases and decisions made when collecting that data, as well as the behaviors and judgements of the groups and individuals from whom the data are collected" (p. 27, emphasis added). Furthermore, the *biases also reflect whatever prejudices* – whether intentional or not – were held by the researchers whose work was referenced when determining what variables or data configurations to use in the algorithms; algorithms only *know* what they are taught, and those who teach them do so using the knowledge *they, themselves,* were taught (Campolo, Sanfilippo, Whittaker, & Crawford, 2017). Campolo et al. (2017) go so far as to suggest that public agencies, including education, "should no longer use 'black box' AI and algorithmic systems" (p. 1), indicating that these systems "should be available for public auditing, testing, review, and subject to accountability standards" (p. 1).

Campolo et al.'s point is well taken, but given the substantial investment many campuses and institutions have already made in analytics, simply abandoning all systems at this point is unrealistic. However, careful reflection by institutional actors must occur (Arnold et al., 2014; Oster et al., 2016). Any biases, at the very least, need to be acknowledged, ideally minimized, and, hopefully, mitigated altogether. Furthermore, institutions contracting with third-party vendors and collaborating with other internal or external entities should discuss the contexts in which data used to train the system was procured, who those data represent, and how decisions to interpret various outputs were made (Zook, et al., 2017). In the end, all parties must work to move more algorithms and predictive models into the light of day.

The more open an institution is about the use of analytics and the data being fed into the system, the more likely the campus community is to buy-in to these efforts and support them broadly. The more accessible and trasparent the algorithms are, particularly with regard to indicating which data points are used to identify various points of concern or risk, the more likely the student is to understand how that concern was raised and the more likely the involved faculty or staff member is to be able to address those concerns with the students.

4. Patience is required by all involved, but especially by those desiring meaningful results.

Learning analytics are not plug-and-play solutions; an institution cannot just purchase and install a system and hope for the best. Further, systems like these take time for full adoption and for results to be identified. Retention is measured in 16- to 32-week chunks. Final grades are only reported every semester, and some behaviors may take multiple terms to be fully realized in students. Campuses must be patient when implementing these systems and be measured in their expectations of success early on in the process. Failure to do so usually results in a quick dismissal of the tool in favor of something else that promises more results more quickly or, potentially, nothing, meaning the initial investment was for naught.

Van Horne, Russell, and Schuh (2015) describe the need to "promote the effective adoption of tools," ensuring that instructors, designers, and educational technologists all understand what a tool will – and will not – do, and, at the same time, laying out reasonable timelines for integration and realizing benefits (p. 7). Further, Davis (1989) notes that the perceived usefulness of a technology is critical to adoption, but if in using technology the usefulness is not realized, the technology will likely be abandoned. Rogers (2000) also points out that absent sufficient support, professional development, and guidance in the use of a technology, instructors will stop using a technology, or worse, never begin using it in the first place. Thus, taking the time to roll out a system purposefully, with careful assessment, can result in a more positive experience for all and allow for implementation alterations to occur along the way rather than as major shifts at the end of a given period.

Additionally, and as Hora (2018) notes and my colleagues and I have written (Arnold et al., 2014; Oster et al, 2016), sufficient capacity must exist, and organizational leadership must be present for institutions to gain the full value of an investment in analytics. Hora, citing Coburn and Turner (2011), further explicates the need to more fully appreciate how faculty and administrators think about data, its usages, and how an organization is equipped to manage these components both during and after implementation. These two aspects are tied to patience. If, at the same time a campus is working to implement analytics technology, capacity does not exist or executive leadership is not actively in support of an effort, then patience will be required to build capacity and/or leadership involvement so that an implementation is successful. Similarly, failure to understand how data is currently perceived or used could stymie the ways in which champions of analytics communicate with various stakeholders. Finally, Knight et al. (2018) illustrate the need for working with stakeholders and not just imposing new technologies or systems on them – another form of patience is required here so that technologies are accepted and integrated into practice, not just acknowledged and ultimately ignored or underused.

These are only some of the issues that may arise. As it may have become apparent to the reader, the issues – both positive and negative – are sometimes universal, but more often are institution and context specific. Whether or not the issue experienced at a given campus is delineated in these pages or not, once identified they should be acknowledged by a large swath of a campus community and worked to be promoted or addressed as appropriate.

How can institutions use learning analytics in a way that enhances both the learning and the teaching process positively and does not punish faculty or students for needing to improve?

The intent and promise of learning analytics implementations are generally positive in nature – identify the students who need assistance in a given class or course of study, then direct them to the appropriate resources in an effort to help them be more successful. In an ideal world, learning analytics can also be used to enhance the teaching environment. This can be done through highlighting areas of information for instructors of large enrollment courses that students failed to grasp so that more time can be spent there. Learning analytics can also be employed to provide instructors with insights into their students *en masse*. For example, an instructor could determine how many students took which prerequisite courses and how they fared in them, to customize content delivery to ensure students would be able to learn the new material well (Bakharia, Corrin, de Barba, Kennedy, Gašević, Mulder, et al., 2016; Gottipati & Shankararaman, 2014; Slim, Heileman, Kozlick, & Abdallah, 2014; Wigdahl, Heileman, Slim, & Abdallah, 2014). Department heads or curriculum coordinators could also utilize analytics to identify which courses best serve as prerequisites or co-requisites or to determine which courses in a sequence are unnecessary or need the greatest amounts of support (Dawson, McWilliam, & Tan, 2008; Slim, Heileman, Al-Douroubi, & Abdallah, 2016).

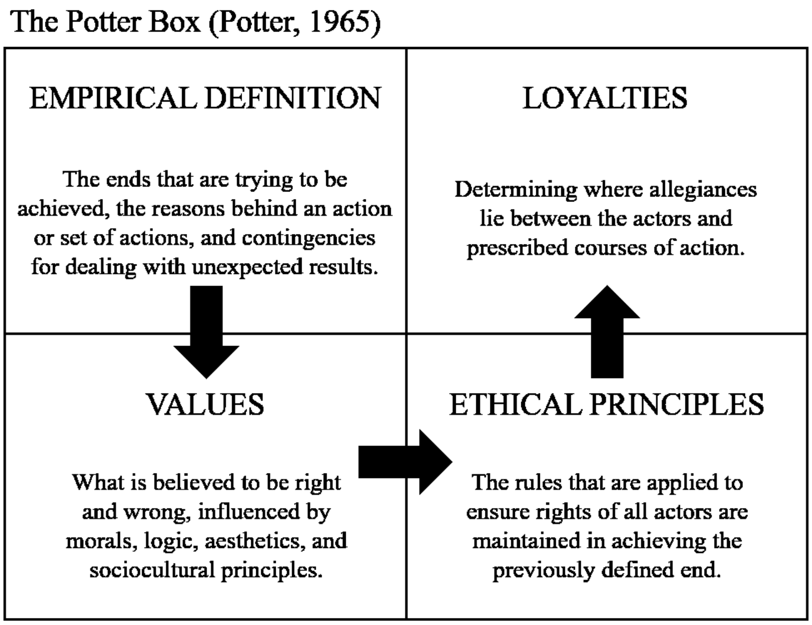
While the use of learning analytics in these ways is admirable, and largely in line with the definitions posed throughout this volume, more nefarious uses can also be imagined and employed. For example, when uploading information about student performance, course numbers, and sometimes instructors' names, are attached. Suppose that a regression equation generated for predicting retention from year to year ultimately includes a negative β that is related to instructor – meaning, the retention of students is directly and negatively related to having had a specific instructor, and were that instructor no longer teaching that course (or at all) then the retention and graduation rates of students in that program might significantly increase. What should be done? Take this example to campuses where faculty unions are in place – how does this new-found exculpatory information affect contract negotiations? How does the interpretation of academic freedom on a given campus or by such groups as the American Association of University Professors allow for this kind of information to be used?

As another example, in response to the environmental pressures institutions face, institutions might look at all previous graduates and only admit students whose data profiles are similar to boost retention and graduation rates. In an effort to reduce student indebtedness, some colleges could opt only to admit those coming from more affluent families based on financial aid figures. Institutions could also use this data to identify instructors whose use of online resources or technologies has led to increased levels of success for historically underserved students, and recognize or reward them for their actions – and allow those instructors to share their practices with the rest of the faculty community.

While not punishment, there are also other scenarios that are both very real and very concerning. Wise discusses in depth that when analytics are "omnipresent" (2014b, slide 31) that the results are either ignored altogether or, more likely, that students will begin to play a numbers game, using analytics to drive them to answers or specific behaviors without actually engaging with the material (Wise, 2014a). This, then, defeats the initial purpose of learning analytics and encouraging higher levels and more effective ways of learning. Beattie, Woodley, and Souter (2014) note that learning analytics can become *creepy*, wherein the data can take on a surveillance tone rather than one of assistance and guidance, in short, where “creepy analytics … carry with them a sometimes hard-to-define undercurrent [beyond] greater engagement and better outcomes” (p. 421).

To avoid this, there are several steps that institutions can take, particularly from an ethical standpoint. Willis, Campbell, and I (2013) entertain a robust discussion of this concept utilizing an ethical model known as the Potter Box (Potter, 1965). The Potter Box (see Fig. 1.) has four components that allow decision makers to fully consider the effect of their potential actions: developing an empirical definition, defining values, determining what ethical principles apply to the situation, and identifying loyalties associated with each of the potential choices to be made. Each aspect is examined in brief below.

Figure 1.



The empirical definitions sought are associated with what an institution is trying to achieve and the reasons for implementing learning analytics in general. In explicitly stating and agreeing to these definitions, institutional actors should clearly understand what is to be expected from an implementation. Further, the possibilities for what may be found are then articulated – along with how that information will be used to improve teaching and learning and not create repercussions for students or instructors. Suppose that in examining course outcomes and linking them with students' course evaluation data, a direct link between the perceptions of and final grades for a specific group of students is identified. An empirical definition that states findings will be used solely for the purposes of improvement and not as a means of punishment will allow for a less confrontational conversation as well as direction for professional development.

At the same time, decisions need to be made about what to do should something outside the original scope of the implementation is identified. For example, a robust analysis of students' performance within a specific course of study to determine how grades in prerequisite courses influence success in later courses also may reveal a larger-than-previously-thought achievement deficit for women. While not initially sought after, this becomes a factor to consider when examining the success of all students and determining courses of action that will address this issue. Having a plan in place for addressing unforeseen findings may help ameliorate future dilemmas.

Through identifying the values associated with what administrators believe to be right and wrong, along with the rights of various actors associated with the implementation and receipt of LA, can be articulated. In this context, these values are tensions, examining how students, faculty, and institutions should act or believe with regard to fulfilling the mission of student success—which may be situation-specific (Prinsloo & Slade, this book), depending on the actor, the location of information, and how success has been defined. Examples include what behaviors students should exhibit to be academically successful, how faculty should adjust their teaching to be effective in the classroom, and what institutions can do with regard to providing resources to faculty, staff, and students to achieve desired retention and graduation rates. In their chapter in this book, Prinsloo and Slade also take care to acknowledge that data often considered nonsensitive can, in fact, be sensitive in a different context. Values, therefore, cannot be static to a situation; instead, they must be flexible enough to both maintain principles espoused by an institution and allow for varying levels of application as a situation or decision point dictates.

A particular point highlighted by Prinsloo and Slade's chapter in this book is the extent to which an institution is able to "decide … whether the data are sensitive or not" *and* the extent to which institutions own students' data or, instead, "have temporary stewardship." This value set is important to consider and navigate. Which party owns or has use of specific data can ultimately dictate what can and cannot happen with that information. The tension created here with regard to the values of ownership and use of data is seemingly addressed by FERPA (1974) in the United States, but issues are created because the legislation does not actually define who owns the data, only who has access to it. In determining the value sets that will be applied to institutional use of data via analytics technology, the reader would do well to strongly consider Prinsloo and Slade's chapter in this volume and their related works. As is the case with this chapter, answers are not fully provided, but a great many points to consider are laid out and can help shape a value-driven way forward.

Ethical principles to apply in the use of learning analytics are going to vary from situation to situation, and thus need to remain flexible – something for which the Potter Box nicely allows. Using the agreed upon empirical definitions, administrators, faculty, and staff can work to identify appropriate models to apply to ensure that the use of analytics is done in an ethical manner. These models can range from the general principles of beneficence (doing the most good for the most people) or maleficence (doing the least amount of harm to the least number of people), or could look to John Stuart Mill's Principle of Utility or Aristolte's Golden Mean for more complex models (Backus & Farraris, 2004; Willis & Pistilli, 2013). Regardless, and as noted by Prinsloo and Slade (2018) in this volume, it is imperative that ethics and consent be fully considered, regardless of the desired or intended outcome associated with the application of learning analytics.

Finally, loyalties need to be considered. In short, this involves looking at each of the three aforementioned principles for the Potter Box and determining if a specific course of action is chosen which party is receiving loyalty from the actor. Individual actors can make these determinations on their own, looking to see if the student, the faculty, or the institution is going to be most directly – and positively – affected by a course of action. Additionally, an institution might also consider building an advisory board or council that has the power to regulate how learning analytics and associated technologies can and will be used on a given campus. By including students, faculty, staff, legal counsel, and key administrators in this group, and making at least a portion of their meetings open to the institutional community creates both openness and transparency. This openness and transparency can help identify where the use of learning analytics are going awry well in advance of there being an actual problem or ethical dilemma to address.

However, even in creating an advisory board, the readiness of a campus to entertain such a body, much less the ability of those appointed to it to competently serve on it, need to be carefully considered. Swenson (2014) encourages administrators to examine who among a campus community has the ability or power to enact change, "to make decisions about the learning analytics model and data [and] legitimize some student knowledge or data and not others" and act accordingly – and in ways that can remedy issues that arise from the study of these aspects (p. 249). Baepler and Murdoch (2010) implore those driving a process to ensure that changes made and actions taken are done so in an effort to effect change for students and faculty, and within learning spaces. The ability to effect change, and to have that change driven by a representative group of campus constituents, will require an institution to have a culture that is ready to allow these things to happen. Readiness, as my colleagues and I have written, requires an appropriate culture to exist (Oster et al., 2016).

Ultimately, how should – not can – data be used, and to what ends?

This is arguably the biggest and most complex question raised within the emerging field of big data at institutions of higher education. As discussed in this chapter and throughout this volume, and as many readers can likely conceive of on their own, there exists the potential for both great good and extraordinary harm within the massive stores of data necessary to fully implement learning analytics. In examining this query, I begin with the second part – what ends are trying to be achieved through the use of big data and learning analytics?

Similar to the pedagogical concept of backwards design (Wiggins & McTighe, 2005), those wishing to achieve a specific end need to begin with that end in mind. Institutions need to consider what the desired outcome they want to achieve is, and then entertain the appropriate methods to work towards that end. Students and faculty should also be consulted as to how they expect analytics to work, what data they believe should be used in the models, and how outcomes might be applied (Dietz-Uhler & Hurn, 2013; Reinholz, Corbo, Bernstein, & Finkelstein, 2018; Whitelock-Wainwright, Gašević, & Tejeiro, 2017). The numerous individual reasons for implementing learning analytics can be rolled up into a few categories: increasing retention/reducing attrition; improving graduation rates; enhancing student learning and classroom outcomes; encouraging students to become integrated both socially and academically with the university community; and, improving the ways faculty and the institution deliver content and provide resources to students. Nearly every conceivable reason for investing in big data processes can fall into one of these categories, and there are few, if any, tools that can meet more than a few of these desires in one software package. Institutions must decide what is of most importance to them and subsequently invest appropriately. Furthermore, the decision to invest cannot be simply because what is important to the organization is that they are, in effect, doing what their peer institutions are doing. The adoption of technology must be end-driven, such that the appropriate and defined application of the technology can actually achieve the desired outcome.

Once an outcome is identified, it is at this point that careful considerations must be made to determine if what is desired is in line with the notions articulated in an institution's mission, vision, and value statements, the thinking of governing boards of regents and trustees, and the expectations of the numerous stakeholders both on campus and within the region and state the institution serves. Often, these may be in conflict with one another. For example, a board of trustees may desire huge increases in first-to-second year retention rates, while the state in which the college lies may be looking for greater access to the institution for its citizens. While both are good and desirable outcomes, the notions of an open access institution and increasing retention rates do not generally coincide. Institutional leaders need to grapple with these groups both independently and jointly in an effort to determine how to work towards both ends without sacrificing ethical principles or a particular group of students or faculty.

So what is to be done? How can a case be made to governing officials at the institutional and state levels that either or neither desire is achievable? Could someone develop a regression model that identifies populations of students not currently enrolled who are likely to be successful and, at the same time, increase retention rates of currently enrolled students? Probably. However, the costs may be not enrolling new students with profiles similar to those currently taking courses at the institution, needing to invest in faculty development or increased faculty numbers in order to meet the demands of the students whose profiles are different from currently enrolled students, or changing the ethos of the institution altogether. None of these are bad, but all carry positive and negative aspects, and barring unlimited resources, cannot be implemented at the same time.

The point of the previous hypothetical situation is to note that the tensions in what are desired by an institution and those it employs, serves, or reports to are often so great that not even data can be used to mitigate the challenges. Does that mean campuses shouldn't invest in big data applications and opportunities? Not at all. It means that institutional leaders need to very carefully choose what they are – and are not – willing to do in the name of a desired outcome.

It is possible to increase student retention rates easily by simply admitting a better academically-qualified group of students. However, this action would result in disenfranchising less academically-qualified students, which disproportionately is likely minority, first-generation, and low income students. Institutions might also focus on increasing the success rates of the students already on campus, but that would at the very least require investment of resources in support programs and associated staff, as well as additional faculty and their development. Both of these approaches are antithetical to the missions of most institutions of higher education, and likely would violate the values and ethical principles administrators espoused in the Potter Box. Consistency between what is desired, value systems, ethical principles, loyalties, and what is eventually implemented must exist; otherwise, the use of data becomes an end unto itself with the potential to effect great harm (Zook, et al., 2017), rather than a mechanism for achieving a specific outcome that has been carefully considered and vetted.

What is to be done – the desired outcomes, the hiring of staff, the changing of an admissions profile – is context specific. What is good for one institution is not the same as what would work for another. Data, even big data, become nothing more than an informant – information that provides greater insight into phenomena or trends, or can help identify the potential outcomes associated with our actions or inactions. *If there is a precise answer as to how data should be used, it is this: to ensure that the decision that will be made about a particular topic is well-informed and rooted in sound analysis, experience, and theory.*

In the absence of good data, or with no data at all, decisions are made based on anecdote (the plural of which is not data) or, sometimes, gut-feel. These decisions are indefensible at best, and rarely achieve the desired outcome. Incomplete data can point administrators in the right direction, but the incompleteness of the data means important considerations or valuable pieces of information are left out, leaving decision makers with a misinformed set of options from which to choose.

References

Adelman, C. (2006). *The toolbox revisited: Paths to degree completion from high school through college*. Washington, DC: U. S. Department of Education. Retrieved from https://www2.ed.gov/rschstat/research/pubs/toolboxrevisit/toolbox.pdf

Arnold, K. E., Lonn, S., & Pistilli, M. D. (2014). An exercise in institutional reflection: The Learning Analytics Readiness Instrument (LARI). In A. Pardo & S. Teasley (Eds.), *Proceedings from the 4th International Conference on Learning Analytics and Knowledge* (pp. 163-167). New York: ACM. DOI: 10.1145/2567574.2567621

Arroway, P., Morgan, G., O'Keefe, M., Yanosky, R. (2016, March). *Learning analytics in higher education* (Research report). Louisville, CO: EDUCAUSE Center for Applied Research. Retrieved from https://library.educause.edu/~/media/files/library/2016/2/ers1504la.pdf

Backus, N., & Farraris, C. (2004). Theory meets practice: Using the Potter Box to teach business communication ethics. Proceedings from the 2004 Association for Business Communication Convention. Cambridge, MA: Association for Business Communication.

Baepler, P., & Murdoch, C. J. (2010) Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning, 4*(2), art. 17. Retrieved from http://digitalcommons.georgiasouthern.edu/ijsotl/vol4/iss2/17

Bakharia, A., Corrin, L., de Barba, P., Kennedy, G., Gašević, D., Mulder, et al. (2016). A conceptual framework linking learning design with learning analytics. In S.

Dawson, H. Drachsler, & C. P. Rosé (Eds.), *Proceedings from the 6th International Conference on Learning Analytics and Knowledge* (pp. 329–338), New York: ACM. doi: 10.1145/2883851.2883944

Beattie, S., Woodley, C., & Souter, K. (2014). Creepy Analytics and Learner Data Rights. In B. Hegarty, J. McDonald, & S.-K. Loke (Eds.), *Rhetoric and Reality: Critical perspectives on educational technology*. Proceedings ASCILITE 2014 (pp. 421-425). Dunedin, New Zealand: ASCILITE. Retrieved from http://www.ascilite.org/conferences/dunedin2014/files/concisepapers/69-Beattie.pdf

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing teaching and learning through educational data mining and learning analytics: An issue brief. Washington, DC: U. S. Department of Education, Office of Educational Technology. Retrieved from https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

Boyer Commission on Educating Undergraduates in the Research University. (1998). *Reinventing undergraduate education: A blueprint for America's research universities.* Stony Brook, NY: The Carnegie Foundation for the Advancement of Teaching. Retrieved from http://files.eric.ed.gov/fulltext/ED424840.pdf

Bracco, K. R., Dadgar, M., Austin, K., Klarin, B., Broek, M., Finkelstein, N., et al. (2014). Core to college evaluation: Exploring the use of multiple measures for placement into college-level courses – Seeking alternatives or improvements to the use of a single standardized test. San Francisco, CA: WestEd. Retrieved from http://eric.ed.gov/?id=ED559630

Campbell, J. P., & Oblinger, D. G. (2007, October). *Academic analytics*. Boulder, CO: EDUCAUSE. Retrieved from http://www.educause.edu/ir/library/pdf/EDU07286B.pdf

Campolo, A., Sanfilippo, M., Whittaker, M., & Crawford, K. (2017). *AI Now 2017 report*. New York: New York University, AI Now Institute. Retrieved from https://assets.contentful.com/8wprhhvnpfc0/1A9c3ZTCZa2KEYM64Wsc2a/8636557c5fb14f2b74b2be64c3ce0c78/_AI_Now_Institute_2017_Report_.pdf

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683-695. doi: 10.1080/13562517.2013.827653

Coburn, C. E., & Turner, E. O. (2011). Research on data use: A framework and analysis. *Measurement: Interdisciplinary Research & Perspective, 9*(4), 173–206. doi: 10.1080/15366367.2011.626729

Dawson, S. P., McWilliam, E. & Tan, J. (2008). Teaching smarter: how mining ICT data can inform and improve learning and teaching practice. Paper presented at the Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (pp. 221-230). Melbourne, Australia: Deakin University.

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions, and behavioral impacts. *International Journal of Man-Machine Studies (38)*, 475-478. doi: 10.1006/imms.1993.1022

DeBerard, M. S., Spielmans, G. I., & Julka, D. C. (2004). Predictors of academic achievement and retention among college freshmen: A longitudinal study. *College Student Journal, 38*, 66–80.

Dietz-Uhler, B., & Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. *Journal of Interactive Online Learning, 12*(1), 17-26.

Family Educational Rights and Privacy Act of 1974, 20 U.S.C. § 1232g (1974).

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *Tech Trends, 59*(1), 64-71. doi: 10.1007/s11528-014-0822-x

Gottipati, S., & Shankararaman, V. (2014). Learning analytics applied to curriculum analysis. Paper presented at Association for Information Systems Special Interest Group for Education. Retrieved from https://aisel.aisnet.org/siged2014/2

Hodara, M., & Lewis, K. (2017). *How well does high school grade point average predict college performance by student urbanicity and timing of college entry?* (REL 2017–250). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Johnson, J. A. (2017). Ethics and justice in learning analytics. In J. Zilvinskis & V. Borden (Eds.), *Learning analytics in higher education* (New Directions for Higher Education No. 179, pp. 77-87). San Francisco: Jossey-Bass. doi: 10.1002/he.20245

McKay, T. (2016). Creating a learning higher education community. New York: Ithaka S+R. doi: 10.18665/sr.283888

Milne, G. R., & Culnan, M. J. (2004). Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices. *Journal of Interactive Marketing, 18*(3), 15-29. doi: 10.1002/dir.20009

Norris, D. (2011). *7 things you should know about first-generation learning analytics*. Louisville, CO: EDUCAUSE Learning Initiative. Available from http://www.educause.edu/ir/library/pdf/ELI7079.pdf

Norris, D., Baer, L., & Offerman, M. (2009). A national agenda for action analytics. Retrieved from http://lindabaer.efoliomn.com/uploads/settinganationalagendaforactionanalytics101509.pdf

Obar, J. A., & Oeldorf-Hirsch, A. (2016). The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. Paper presented at TPRC 44: The 44th Research Conference on Communication, Information and Internet Policy 2016. Retrieved from https://ssrn.com/abstract=2457465. doi: 10.2139/ssrn.2757465

Online Student Support Services. (2016). Learning analytics and you [Webpage]. Retreived from http://www2.open.ac.uk/students/help/learning-analytics-and-you

Oster, M., Lonn, S., Pistilli, M. D., Brown, M. G. (2016). The Learning Analytics Readiness Instrument. In S. Dawson, H. Drachsler, & C. P. Rosé (Eds.), *Proceedings from the 6th International Conference on Learning Analytics and Knowledge* (pp. 173-182), New York: ACM. doi: 10.1145/2883851.2883925

Pistilli, M. D., Willis, J. E., & Campbell, J. P. (2014). Analytics from an institutional lens: Definition, theory, design and impact. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 79-102). New York: Springer Books, Computer-Supported Collaborative Learning series. doi: 10.1007/978-1-4614-3305-7

Potter, R. B. (1965). *The structure of certain American Christian responses to the nuclear dilemma, 1958-1963*. Unpublished doctoral thesis, Harvard University. Retrieved from https://dash.harvard.edu/handle/1/32749952

Rogers, P. L. (2000). Barriers to adopting emerging technologies in education. *Journal of Educational Computing Research, 22*(4), 455-472. doi: 10.2190/4UJE-B6VW-A30N-MCE5

Slim, A., Heileman, G. L., Al-Doroubi, W., & Abdallah, C. T. (2016). The impact of course enrollment sequences on student success. Paper presented at the IEEE 30th International Conference on Advanced Information Networking and Applications. Retrieved from http://ieeexplore.ieee.org/document/7474070/. doi: 10.1109/AINA.2016.140

Slim, A., Heileman, G. L., Kozlick, J., & Abdallah, C. T. (2014). Predicting student success based on prior performance. Paper presented at the IEEE Syumposium on Computational Intelligence and Data Mining. Retrieved from http://ieeexplore.ieee.org/abstract/document/7008697. doi: 10.1109/CIDM.2014.7008697

Society of Learning Analytics Research. (2012). About [Webpage]. Retreived from http://www.solaresearch.org/mission/about/

Steele, C. M. & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*(5), 797-811. doi: 10.1037/0022-3514.69.5.797

Stumpf, H., & Stanley, J. C. (2002). Group data on high school grade point averages and scores on academic aptitude tests as predictors of institutional graduation rates. *Educational and Psychological Measurement, 62*, 1042–1052. doi: 10.1177/0013164402238091

Swenson, J. (2014). Establishing an ethical literacy for learning analytics. In A. Pardo & S. Teasley (Eds.), *Proceedings from the 4th International Conference on Learning Analytics and Knowledge* (pp. 264-250). New York: ACM. doi: 10.1145/2567574.2567613

U. S. Department of Education. (2006*). A test of leadership: Charting the future of U. S. higher education*. *A report of the commission appointed by Secretary of Education Margaret Spellings.* Washington, DC: Author. Retrieved from https://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

Van Horne, S., Russell, J., & Schuh, K. L. (2015). Assessment with e-textbook analytics. Research bulletin. Louisville, CO: EDUCAUSE Center for Analysis and Research. Retreived from https://library.educause.edu/~/media/Files/Library/2015/2/erb1501-pdf.pdf

Waldman, A. E. (in press). A statistical analysis of privacy policy design. *Notre Dame Law Review Online (93)*. Retrieved from https://ssrn.com/abstract=2940505

Whitelock-Wainwright, A., Gašević, D., & Tejeiro, R. (2017). What do students want? Towards an instrument for students' evaluation of quality of learning analytics services. In X. Ochoa, I. Molenaar, & S. Dawson (Eds.), *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 368-372), New York: ACM. doi: 10.1145/3027385.3027419

Wigdahl, J., Heileman, G. L., Slim, A., & Abdallah, C. T. (2014). Curricular efficiency: What role does it play in student success? Paper presented at the 121st ASEE Annual Conference & Exposition. Retrieved from http://www.asee.org/file_server/papers/attachment/file/0004/3560/CurricularEfficiency-Final.pdf

Wiggins, G., & McTighe, J. (2005). *Understanding by design* (2nd Ed.). Alexandria, VA: ACSD.

Willis, III, J. W., Campbell, J. P., & Pistilli, M. D. (2013). Ethics, big data, and analytics: A model for application. *EDUCAUSE Review Online.* Retrieved from http://er.educause.edu/articles/2013/5/ethics-big-data-and-analytics-a-model-for-application

Wise, A. F. (2014a). Designing pedagogical interventions to support student use of learning analytics. In S. Teasley & A. Pardo (Eds.), Proceedings of the 4th International Conference on Learning Analytics and Knowledge (pp. 203-211), New York: ACM. doi: 10.1145/2567574.2567588

Wise, A. F. (2014b). Designing pedagogical interventions to support student use of learning analytics [PowerPoint slides]. Retrieved from https://www.slideshare.net/alywise/designing-pedagogical-interventions-to-support-student-use-of-learning-analytics

Zook M., Barocas S., boyd d., Crawford K., Keller E., Gangadharan S. P., et al. (2017) Ten simple rules for responsible big data research. *PLoS Computational Biology 13*(3): e1005399. doi: 10.1371/journal.pcbi.1005399