

5-6-2013

Ethics, Big Data, and Analytics: A Model for Application.

James E. Willis III

Purdue University, jewillis@purdue.edu

Follow this and additional works at: <http://docs.lib.purdue.edu/idcpubs>

Recommended Citation

Willis, James E. III, "Ethics, Big Data, and Analytics: A Model for Application." (2013). *Instructional Development Center Publications*. Paper 1.
<http://docs.lib.purdue.edu/idcpubs/1>

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.

Ethics, Big Data, and Analytics: A Model for Application

Published on Monday, May 6, 2013 | 0 Comments

The use of **big data and analytics to predict student success** presents unique **ethical questions** for higher education **administrators** relating to the nature of knowledge; in education, **"to know" entails an obligation to act** on behalf of the student. The **Potter Box framework** can help administrators address these questions and **provide a framework for action**.



Big data and analytics, which "marries large data sets, statistical techniques and predictive modeling [to mine] institutional data to produce 'actionable intelligence,'"¹ present big questions to those of us in higher education. Systems now exist that link students' demographic data with data on past and current educational performance, engagement with online course materials, or in-class participation levels to predict, with startling accuracy, specific outcomes such as final grades within a course. When applied to collegiate administration, this raises several ethical questions, including:

- What, specifically, is the role of big data in education?
- How can big data enrich the student experience?
- Is it possible to use big data to increase retention?
- To what extent can big data contribute to successful outcomes?

More specifically, we must ask what it means to "know" with predictive analytics. Furthermore, once an administration "knows" something about student performance, what ethical obligations follow?

To contextualize some of the problems, David Kay, Naomi Korn, and Charles Oppenheim's recent work discusses the "obligation to act" imposed by the college administration's "knowledge and the consequent challenges of student and employee performance management."² This obligation is related to "underlying issues of learning management, including social and performance engineering."³ Their study charts specific

issues of educational analytics in terms of the Nuremberg, Belmont, and Helsinki codes; furthermore, it asks the very pertinent question of whether or not people are willing "to trade privacy for other benefits within the new social economy."⁴ Marie Bienkowski, Minguy Feng, and Barbara Means note that the possession of longitudinal big data could "result in disclosure [that] may be hard to foresee"⁵ — de-identified information can be realigned to tease out more valuable information. Thus, the obligation of knowing extends from considering students' rights to knowledge that enables "social engineering" to the institutional obligation to protect sensitive data sets.

Contextual Ethics: Collegiate Administration

As Samuel Thompson defines it,⁶ ethics is concerned with the two basic practical problems of human life:

- What is worth seeking — that is, what ends or goals of life are good?
- What individuals are responsible for — that is, what duties should they recognize and attempt to fulfill?

This definition's ramifications are pertinent to higher education, particularly in the relations among students, faculty, and administrators.

The question of what is worth seeking in higher education is influenced by many factors; however, we believe that everyone involved in the educational enterprise views academic success for students as a worthy goal. Determining who is responsible for academic success, however, is more difficult. Is it the student, the professor, the administration of the college or university, or all of these? Does one individual or group hold more responsibility for academic success than another? Here, the question is directly related to questions about collegiate administration — specifically, the ethical questions that arise from the *obligation of knowing*.

Raising the Ethical Question: The Power of Knowing through Analytics

Technology-based instructional tools automatically collect a wide array of data on student usage. Analytic techniques utilizing this data provide new insights into which students are performing well and which students need additional help.

One of the first full-scale analytics projects, Purdue University's Signals predicts which students are at risk of doing poorly using data such as student demographics and academic history, engagement with online resources, and current performance in a given course.⁷ Based on the prediction, students receive a red, yellow, or green light; the red light indicates "stop and get help," yellow is "caution you are falling behind," and green is "keep on going." With each light, students receive instructor feedback on their performance and where to go next.

Purdue University's work has led to significant gains in overall retention. For example, the four-year retention rate for Purdue students in the 2007 cohort without a Signals course was 69.40 percent ($n = 5,134$ students); students in the same cohort with at least one Signals course had a retention rate of 87.42 percent ($n = 1,518$ students). Even more encouragingly, students with two or more Signals courses had a retention rate of 93.24 percent ($n = 207$ students).⁸ The retention numbers skyrocketed for Signals courses, yet the incoming student SAT scores for students in two or more Signal courses was more than 50 points lower than for the no-courses group.⁹ What this tells us is that students who are less prepared for college — as measured *solely* by standardized test score — are retained by and graduated from Purdue at higher rates than their better-prepared peers after having one or more courses in which Signals was used. In short, by receiving regular, actionable feedback on their academic performance, students were able to alter their behaviors in a way that resulted in stronger course performance, leading to enhanced academic performance over time.¹⁰ Student perceptions of Signals were equally encouraging: 89 percent of students using Signals had a positive experience, and 74 percent said "their motivation was positively affected" by it.¹¹ The Signals program's success is directly attributable to using big data to offer direct feedback.

The use of analytics to predict academic success in Signals and other programs compels many institutions to

examine the ethical issues of teaching, including accountability and the distribution of resources. With access to these predictive formulas, faculty members, students, and institutions must confront their responsibilities related to academic success and retention, elevating these key issues from a "general awareness" to a quantified value.

The ethical issues related to academic success and retention are probably best examined as "what to do" with the new knowledge, rather than "what is the right answer" to the treatment of big data sets.¹² That is, the focus is pragmatic, examining how to use information about the likelihood of academic success in a meaningful and effective manner. Likewise, the question of faculty member, student, and institutional responsibility should be examined in terms of: what the options are, where we are at the time, and what we can do. "Knowing" the likelihood of academic success forces each party to examine the issues; participation is unavoidable. However, using academic data does not trap faculty members into a particular approach; the question of how to use data has many answers and must be weighed against higher education's changing environment and the shifting expectations of key stakeholders.

The Ethical Dilemma of Analytics: The Potter Box

Using analytics on student data entails an interesting set of ethical issues. While not exhaustive, the following list exemplifies the types of questions institutions must address when using big data:

- Does the college administration let students know their academic behaviors are being tracked?
- What and how much information should be provided to the student?
- How much information does the institution give faculty members? Does the institution provide a calculated probability of academic success or just a classification of success (e.g., above average, average, below average)?
- How should the faculty member react to the data? Should the faculty member contact the student? Will the data influence perceptions of the student and the grading of assignments?
- What amount of resources should the institution invest in students who are unlikely to succeed in a course?
- What obligation does the student have to seek assistance?

To take a deeper look at these questions, we use *The Potter Box*,¹³ a popular ethical model in business communications. The Potter Box is dynamic and allows for multiple and varied analyses; this is especially useful in situations where big data might be used to assess the direction of an individual student's academic path. Although other ethical models could be applied here, The Potter Box provides the most analytically powerful yet adaptable model, and thus might appeal to many levels of collegiate administration.

Originally developed from his doctoral work on responses to nuclear proliferation in the early 1960s, Ralph B. Potter, Jr., devised his box as a series of concepts that help its users outline a problem, sort through conflicting ideas, appeal to established ethical principles, and devise a judgment.¹⁴ Although there are slight variations in how the model's boxes are defined, we use Rod Carveth's action-based labels¹⁵:

- Provide an empirical definition
- Identify values
- Appeal to ethical principles
- Choose loyalties

Carveth explains that the boxes "are a linked system, a circle, or an organic whole and not a random set of

isolated questions." From here, The Potter Box facilitates "making a particular judgment or policy, and finally, providing feedback" to those asking the question. Potter argues that, "any sustained argument, waged by alert, persistent interlocutors, would have, eventually, to deal with each of the four types of questions I had isolated."¹⁶ We now offer a snapshot of how The Potter Box can be applied to a specific set of questions; as the following shows, the model's versatility and adaptability make it a useful tool in this context.

Providing an Empirical Definition

Ethics are highly nuanced, and if they are relativistic, "then there are no common, overarching moral truths to be known."¹⁷ One of the strengths of The Potter Box is that it helps us clearly define the possible "facts" in a given situation. This requires that we try to think as objectively as possible.¹⁸ For this project's purpose, we assume the following facts.

- Data on students' academic behaviors are collected automatically by various technology-based instructional tools. Students and faculty members are largely unaware that the online tools are collecting the data.
- The data collected can be compiled and analyzed to determine if there is a correlation between the student behaviors displayed within the course management system and academic success, particularly when combined with other information on the student, including application information and previous course-performance records. Technology can provide greater institutional analysis than was conceivable in previous years.
- Based on the current research, the data have been found to be highly predictive of student success, as defined by increased grades in a course and higher levels of persistence and graduation rates.¹⁹ Based on initial success, early intervention programs can be initiated based on the data.
- Over time, additional data sources will become available for more sophisticated analysis.
- The emphasis on student academic success and retention continues to increase due to pressure from students, parents, legislators, and the community as each reacts to the increasing cost of higher education and the necessity of higher education for individuals to secure long-term employment.

Identifying Values

Once we have empirical definitions, it is possible to state and compare the merits of various values that represent our notions of rights, beliefs, and right and wrong, as well as "questions of facts and values."²⁰ From a collegiate administration standpoint, the tension in values is important here:

- Students *should* be provided with prompt feedback on their educational progress.²¹
- Students *should* actively seek help.²²
- Students are independent young adults capable of making decisions about their education, careers, and values. They *should* be held responsible for these decisions.²³
- Faculty members *should* take advantage of the tools available to improve learning within the classroom.²⁴
- Faculty members *should* foster and promote reflection and lifelong learning.²⁵
- Faculty members *should* actively promote the resources available to the student to promote academic success.²⁶

- Institutions *should* admit only students who are qualified to pursue an academic degree.²⁷
- Institutions *should* provide faculty and students with the resources necessary to meet the educational mission.²⁸

Appealing to Ethical Principles

Entire ethical systems can be rather vague and unapproachable, and thus impracticable. What really gives an ethical system teeth is a set of principles. What makes The Potter Box dynamic in the realm of big data and analytics is the possibility of concurrently examining multiple principles. Many educators, administrators, and researchers have been drawn to the use of educational analytics to provide real-time, actionable intelligence to alter student behaviors.²⁹ Any discussion of ethical principles must consider what is actionable for the student. The following are some helpful ethical principles³⁰:

- **Aristotle's Golden Mean: "The moral virtue is the appropriate location between two extremes."** To establish opposite ends of an educational spectrum, we chose the categories of "deficiency" and "excess." Given these categories, we can identify certain quantifiable, actionable student behaviors that suggest a golden mean. As Table 1 shows, in this model analytics and big data can be used to identify the extremes of student behaviors and thus suggest a move to the middle. Although this list is by no means exhaustive, it shows how identifying extremes suggests behavior changes that might compel students to move to the middle ground between deficiency and excess.

Table 1. Extremes of Student Behavior Suggest a Golden Mean

Example Behavior	Deficiency	Excess
Performance	Under-confidence	Over-confidence
Analytical thinking	Uncritical	Unable to synthesize (all knowledge remains fractured)
Attitude	Apathy/lack of effort	Unable to step away from education (obsessive demand for success)

- **Immanuel Kant's Categorical Imperative: "Act on the maxim that you wish to have become a universal law."** A college's administration has a unique duty once its chosen algorithms have demonstrated their effectiveness in predicting outcomes and providing actionable knowledge. At this point, the administration has a relational duty to make decisions in a way that is devoid of personal motivations. In decision-making based on analysis of big data, college administrations must consider carefully the implications of universally applying what is "known" through algorithmic logic.
- **John Stuart Mill's Principle of Utility: "Seek the greatest happiness for the greatest number."** College administration ought to consider what utilitarian "happiness" is in terms of student success. Big data can provide unique insight into what we might call "happiness," provided it is framed in terms of learning objectives and grade outcomes. Both of these are quantifiable and, thus, demonstrable of what really works; big data can provide the matrix of successful behaviors that can contribute to outcomes indicative of real learning.
- **John Rawls' Veil of Ignorance: "Justice emerges when negotiations are without social differentiation."** Using big data in analytics might create an ethical dilemma with the Veil of Ignorance, or a method in which impartial judgment is required. Although data can be de-identified, it is often analyzed in terms of inferential statistics that operate under categories such as gender, ethnicity, and income. Aggregated data can provide incredible knowledge to a college administration, but the model's risk is that the framework of categories can artificially impact decisions. Further, aggregated data eliminates outliers, which also can be

problematic. Is it possible or beneficial to perform deep analysis while maintaining the Veil of Ignorance over differentiations? Being aware of potential biases that might emerge when pinpointing particular data can help guide meaningful discussions.

Choosing Loyalties

Impending ethical decisions compel an honest evaluation of critical loyalties, which can be identified with the help of various questions. These questions include taking stock of each of the above principles and asking, "If I base my action in this situation on this principle, to whom am I being loyal?"³¹ Also, considering other possible loyalties — and, importantly, the person or people associated therein — better defines which loyalties should be affirmed or denied.³²

Loyalty questions that a collegiate administration might consider include:

- Who is affected by the analysis or application of big data, and how should they be affected by it?
- Once a college's administration has the tools to "know" with statistical significance those who might be in jeopardy of failing, who is compelled to act on that knowledge?
- What action is appropriate based on the information learned as a result of the analysis?
- Who is responsible when a predictive analytic is incorrect?

Responsibilities and Recommendations

The Potter Box lets us analyze various ethical questions related to "knowing" the likelihood of academic success based on student behaviors within technology-based instructional tools. Given this, we propose the following responsibilities:

- The institution is responsible for developing, refining, and using the massive amount of data it collects to improve student success and retention, as well as for providing the tools and information necessary to promote student academic success and retention.
- The institution is responsible for providing students and faculty members with the training and support necessary to use the tools in the most effective manner. It further is responsible for providing students with excellent instructional opportunities, student advising, and a supportive learning environment, as well as for providing faculty members with tools that allow them to deliver timely feedback to students on their progress within their courses.
- The institution is responsible for providing a campus climate that is both attractive and engaging and that enhances the likelihood that students will connect with faculty and other students, and for recognizing and rewarding faculty and staff who are committed to student academic success and retention.

Institutions must determine their own individual goals as they provide resources to help students succeed. Regardless of one's opinion of the ethical obligations, it is evident that, in one form or another, analytics will continue to provide opportunities to begin a dialogue on the distinct yet interrelated roles of faculty members, students, and institutions in the learning process.

Adjusting the language in orientation programs and publications to match the institutional philosophy on student self-directedness will contribute to this dialogue. The extent to which students are expected to control their own college experience should not be vague. For example, if the role of faculty members is not to answer every question, but rather to teach students how to own their own learning, institutions ought to ensure that this is

stated in all communication about academic success. Similarly, students should not be given the impression that the academic environment on campus is completely controlled if it is not; students should know that resources are available to help with academic difficulties, but that the ultimate choice on whether or not to use those resources is their decision and responsibility. It is important, however, for institutions to use big data and analytics to help appropriately guide students.

Ethical Relativism?

When applying ethical inquiry to any new method or practice, it can easily shift into ethical relativism, where no fixed principles universally apply to any situation that may arise. Using big data in statistical modeling indeed marks a new method to provide better education to students. It is beyond this article's scope to suggest whether big data use is better examined as a universal construct or as a relativistic enterprise, but the goal is to suggest that more work is needed to pull out the ethical implications in the obligation of knowing.

Further, the obligation of knowing might lend itself to new ethical horizons. The above principles encapsulate highly contextualized ways of administrative knowing well, but they do not address the entirety of goals, resources, and outcomes of a given collegiate administration. Simply put, practical ethics operate on one part of the equation and theoretical ethics operate on the other; the intersection of the two is where we can take action and have the greatest effect. In the end, are these not new ways of knowing? Futuristic models of student outcomes based on existing student data begs what it means to know within a complex web of administrative decision-making.

Conclusion

The intersection of big data, analytics, college administration, and ethical reflection is best examined within the obligation of knowing paradigm. The real problem here is that statistical probability within a matrix of academic prediction can have massive consequences for institutions and individual students alike. There simply has not been enough ethical reflection on these issues to date. Like other forms of technology, the development of software platforms and predictive analytics quickly outpaces the larger questions of ethical ramifications and the time it takes to consider even the smallest implications.

Current analytics projects illustrate the potential ethical questions that could arise from using academic data to predict student success in higher education. Eventually, institutions will find themselves in the awkward position of trying to balance faculty expectations, various federal privacy laws, and the institution's own philosophy of student development. It is therefore critical that institutions understand the dynamic nature of academic success and retention, provide an environment for open dialogue, and develop practices and policies to address these issues.

Notes

1. John P. Campbell, Peter B. DeBlois, and Diana G. Oblinger, "[Academic Analytics: A New Tool for a New Era](#)," *EDUCAUSE Review*, vol. 42, no. 4, 2007.
2. David Kay, Naomi Korn, and Charles Oppenheim, "[Legal, Risk and Ethical Aspects of Analytics in Higher Education](#)," *Center for Educational Technology & Interoperability Standards*, vol. 1, no. 6, 2012.
3. *Ibid.*, p. 5.
4. *Ibid.*, p. 25.
5. Marie Bienkowski, Minguy Feng, and Barbara Means, [Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief](#), Office of Educational Technology, U.S. Department of Education, Oct. 2012, p. 42.
6. Samuel M. Thompson, *The Nature of Philosophy: An Introduction*, Holt, Rinehart, and Winston, New York,

1961, p. 304.

7. Kimberly E. Arnold, Zeynep Tanes, and Abigail Selzer King, "[Administrative Perceptions of Data-Mining Software Signals: Promoting Student Success and Retention](#)," *The Journal of Academic Administration in Higher Education*, 2010, pp. 29–40.
8. Kimberly E. Arnold and Matthew D. Pistilli, "[Course Signals at Purdue: Using Learning Analytics to Increase Student Success](#)," *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, ACM Press, 2012, p. 267–270.
9. Dale Whittaker, "Data to Decision, Student to Institution: Evidence-Based Improvement in Universities," keynote address, *Society of Learning Analytics Research Flare Conference*, Purdue University, 2012.
10. Ibid.
11. Arnold and Pistilli, "[Course Signals at Purdue](#)," p. 269.
12. Gregory Pappas, "The Reconstruction of Dewey's Ethics," in *In Dewey's Wake: Unfinished Work of Pragmatic Reconstruction*, William J. Gavin, ed., State University of New York Press, 2003.
13. Nick Backus and Claire Ferraris, "[Theory Meets Practice: Using the Potter Box to Teach Business Communication Ethics](#)," *Association for Business Communication*, 2004, p. 222; and Ralph B. Potter, *The Structure of Certain American Christian Responses to the Nuclear Dilemma, 1958–1963*, doctoral dissertation, Harvard University, 1965.
14. Rod Carveth, "[The Ethics of Faking Reviews about Your Competition](#)," *Proceedings of the 76th Annual Convention of the Association for Business Communication*, Association for Business Communication, 2011, pp. 4–5.
15. Carveth, "[The Ethics of Faking Reviews about Your Competition](#)," p. 6.
16. Backus and Ferraris, "[Theory Meets Practice](#)," p. 224.
17. R. Scott Smith, *Virtue Ethics and Moral Knowledge: Philosophy of Language After MacIntyre and Hauerwas*, Ashgate Publishing Limited, Burlington, 2003, p. 195.
18. Carveth, "[The Ethics of Faking Reviews about Your Competition](#)," p. 6.
19. Campbell, DeBlois, and Oblinger, "[Academic Analytics](#)," p. 44; Arnold, Tanes, and King, "[Administrative Perceptions of Data-Mining Software Signals](#)," pp. 32–33; Cali M. Davis, J. Michael Hardin, Tom Bohannon, and Jerry Oglesby, "Data Mining Applications in Higher Education," in *Data Mining Methods and Applications*, Kenneth D. Lawrence, Stephan Kudyba, and Ronald K. Limberg, eds. (New York: Auerbach Publications, Imprint of Taylor & Francis Group, 2008), pp. 132–140; and Arnold and Pistilli, "[Course Signals at Purdue](#)."
20. Carveth, "[The Ethics of Faking Reviews about Your Competition](#)," p. 7.
21. Arthur W. Chickering and Zelda F. Gamson, "[Seven Principles for Good Practice in Undergraduate Education](#)," *AAHE Bulletin*, vol. 39, no. 7, 1987.
22. Stuart A. Karabenick and John R. Knapp, "Relationship of Academic Help Seeking to the Use of Learning Strategies and Other Instrumental Achievement Behavior in College Students," *Journal of Educational Psychology*, vol. 83, no. 2, 1991.
23. Todd M. Davis and Patricia H. Murrell, "A Structural Model of Perceived Academic, Personal and Vocational Gains Related to College Student Responsibility," *Research in Higher Education*, vol. 34, no. 3, 1993; and David A. Kolb, *Experiential Learning: Experience as the Source of Learning and Development*, Prentice-Hall,

New Jersey, 1984.

24. Nancy Hyland and Jeannine Kranzow, "[Faculty and Student Views of Using Digital Tools to Enhance Self-directed Learning and Critical Thinking](#)," *International Journal of Self-Directed Learning*, vol. 8, no. 2, 2011.
25. Susan Lord, John C. Chen, Katharyn Nottis, Candice Stefanou, Michael Prince, and Jonathan Stolk, "[Role of Faculty in Promoting Lifelong Learning: Characterizing Classroom Environments](#)," in *Proceedings of Education Engineering (EDUCON)*, IEEE Press, 2010.
26. Arthur W. Chickering and Stephen C. Ehrmann, "[Implementing the Seven Principles: Technology as Lever](#)," *AAHE Bulletin*, vol. 49, no. 2, 1996.
27. Roger L. Geiger, *Knowledge and Money: Research Universities and the Paradox of the Marketplace*, Stanford University Press, 2004.
28. Vincent Tinto, "[Taking Student Success Seriously: Rethinking the First Year of College](#)," in *Ninth Annual Intersession Academic Affairs Forum*, California State University, Fullerton, 2005.
29. Campbell, DeBlois, and Oblinger, "[Academic Analytics](#)," p. 40.
30. Backus and Ferraris, "[Theory Meets Practice](#)," p. 226.
31. Backus and Ferraris, "[Theory Meets Practice](#)," p. 227.
32. Backus and Ferraris, "[Theory Meets Practice](#)," pp. 226–227.

© 2013 James E. Willis, III, John P. Campbell, and Matthew D. Pistilli. The text of this *EDUCAUSE Review Online* article is licensed under the [Creative Commons Attribution-NonCommercial-NoDerivs 3.0 license](#).

Tags from the EDUCAUSE Library

- [Analytics](#)
- [Data Privacy](#)
- [Student Retention](#)
- [Student Success](#)
- [Big Data](#)

Tags from the Community

- [Ethics](#)

[Log in to comment](#)