The Ethical Use of Student Data and Analytics

The Reinvention Center
Student Success/Learning Analytics Specialized Network

Laura Jensen PhD, Colorado State
Vanessa Roof PhD, University of Nebraska - Lincoln

A rapidly growing trend among higher education institutions is to increase the abundance and availability of student data. Included are demographic, family of origin, postsecondary readiness, academic performance, course pedagogy, curriculum design, digital objects, and a host of other factors that can provide insight into the individual, their environment, and the related interactions. Institutions maintain data from Admissions, Student Financial Aid, Business and Financial Services, Academic Affairs, Student Affairs, and Career Services. In large part, thanks to card swipe and learning management systems, some institutions are more aware of how students spend their time and money and how actively they engage with course content. Additionally, data about outcomes after graduation such as employment, wages, volunteerism, military service etc. are collected and maintained.

There are many reasons this breadth and depth of data is increasingly important. Data allow institutions to teach and assist students, empirically demonstrate value added, show responsible stewardship of public funds, comply with state and federal reporting mandates (including those related to Title IV), document compliance with regional accreditation standards and recruit new students. These data can also provide new insights into student behaviors related to learning. They have the potential to shed light on which curriculum designs, instructional delivery, and academic support programs most effectively promote student success. Such insights can play a decisive role in helping at-risk students to learn and eventually graduate. Thus, institutions have an ethical responsibility to collect and use these data on behalf of students.

However, while the surge in available data is reshaping the foundations of higher education, it increases the risk for "analysis paralysis," reckless collection/storage, and improper use of data. In 2014, legal and ethical scholars assembled at the Asilomar Convention for Learning Research in Higher Education and developed a framework to inform the use of data and technology in learning research ("The Asilomar Convention" 2015). Two fundamental principles materialized regarding the use of student data and results related to analytics in higher education. The first principle was that learning research should be for the purpose of improving higher education. The second principle was that data and results should be shared only with individuals and organizations committed to shared ethics. These principles offer important guidance, but further leadership on the ethical use of student data is sorely needed.

As a consortium of 67 of the nation's research universities, The Reinvention Center is in a position to help provide such leadership. The current white paper is a product of the Reinvention Center's Student Success/Learning Analytics Specialized Network summer 2015 meeting, which identified key challenges related to collecting, analyzing and sharing data while it facilitated careful consideration of competing interests, possible solutions and use cases.

Student Success/Learning Analytics Specialized Network Summer 2015 Meeting Results

During the summer session, participating institutions were surveyed on the landscape of analytics investments across each campus. Some consistent themes were revealed, suggesting a collective mindset. Major themes were as follows:

- Campuses are still discovering who governs analytics and are just beginning to form a standing data governance group to wrestle with issues of access, architecture, ethics and appropriate use of data.
- Due to the expansive amount of available data and complex (sometimes siloed) systems of record, many units across campus need to be involved in driving decisions, managing systems (homegrown and third party) and integrating data in meaningful ways. Efforts must be highly collaborative and involve Information Technology leadership and institutional legal counsel.
- Campuses have many initiatives to promote student success. These initiatives are increasingly data-informed and are now in evaluative stages to determine which efforts provide the biggest return on investment.
- Campuses have identified key courses that impact student success. While analytics can help identify which courses impact successful trajectories, many theories still remain as to the reasons these courses hold significant impact. Learning analytics will be an important foundation for efforts in curricular and pedagogical redesign related to these pivotal courses.
- Campuses are concerned with how to sustain the focus of analytics on continuous quality improvement instead of on compliance (such as accreditation).
- It is understood that the greatest improvements to student learning are attained through collaborations between faculty, support staff and students.
- The integration of high impact practices into programs of study or even individual courses has the potential to deepen student learning by providing opportunities for students to transfer knowledge and skills to novel situations. Data collection related to high impact practices may well include co-curricular activities. Therefore, integration between curricular and co-curricular data systems is important to a holistic understanding of student learning.
- Students need to be involved in discussions about their personal and academic success.
- Students should be informed of which data are being collected, why it is important and how the collection accomplished. Institutions will be most successful when they are clear and direct with students, detailing academic options and the rationale behind interventions.

The remainder of this white paper elaborates on these themes and incorporates input from many of the Reinvention Center institutions.

Data Collection, Transparency and IRB

Primary to the discussion of data collection is the issue of transparency. According to Family Educational Rights and Privacy Act (FERPA) legislation and regulations, written consent is not required in order for institutions (or organizations acting on the institution's behalf) to use student data in research to improve instruction or for other legitimate educational interests. State privacy laws may be more restrictive. Regardless, institutions do have a responsibility to inform students, other stakeholders, and the general public about the data that is being collected, as well as the potential risks and benefits involved. Risks would include a possible, although unlikely, data breach. Students should be made aware of the institution's policies and practices to mitigate this risk. They should also be made aware of the potential benefits of data analysis/research including personalized transition support, better academic advising, stronger curricula and instructional approaches, appropriate levels of academic support and institutional

improvement for the benefit of future students. Further, students should be reminded of their right to inspect and review their educational records at their discretion.

Also relevant to the discussion of data collection is the involvement of the Institutional Review Board/Independent Ethics Committee. The IRB/ICE will decide if analytics-based projects require their continuing oversight or if projects qualify for exemption. Nevertheless, methodologies should be transparent to students, parents, and faculty/staff. The Code of Federal Regulations (45 CFR 46.101) allows exemption from IRB review for research activities in which the <u>only</u> involvement of human subjects will be in one or more of the following categories:

(1) Research conducted in established or commonly accepted educational settings, involving normal educational practices, such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

(2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, or reputation.

(3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if: (i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) require(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.

(4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subjects.

(5) Research and demonstration projects which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) Public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.

(6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the Food and Drug

Administration or approved by the Environmental Protection Agency or the Food Safety and Inspection Service of the U.S. Department of Agriculture.

Data Governance

Data governance provides a crucial role in monitoring the ethical use of student data for analytical investigation. A governing board provides authority over data architecture, access and training while also developing policies and procedures that effectively protect student data and guide faculty/staff on ethical and permissible uses. A standing committee with membership from major institutional functional areas should be established to ensure this ongoing oversight. These members should agree, fundamentally, that data from functional areas are not "owned" by that area and are, instead, institutionally "owned" and should be used for the greatest good. Members should also be expected to:

1) Possess a thorough understanding of the nuances of their data
2) Possess a familiarity with the applicability of relevant local, state, and federal laws/regulations
3) Be responsible for training on the appropriate understanding, collection, use and storage of data
4) Be responsible for the accuracy of the data generated within their functional area
5) Recommend levels of access to specific data elements and user roles to help safeguard against over exposure
6) Actively engage in discussions about specific use cases on campus

Confidentiality

Confidentiality of the educational record is critical, however, student success/learning analytics create circumstances where boundaries can be uncertain. FERPA details specific cases where disclosure of personally identifiable information is appropriate. One such case is where safety or health of students, faculty, or staff might be in jeopardy. As data from student affairs (advisors, RAs, tutoring, etc.) is increasingly used in analytics, these specific circumstances may arise. Therefore, all those using data should remember that, in this circumstance, confidentiality must be broken and the campus' threat assessment team or police should be notified. Outside of these health and safety emergency circumstances, every effort should be made to maintain confidentiality. All data users should be required to adhere to standard security policies and audits should be periodically conducted.

Appropriate Uses of Data

A utilitarian perspective often guides ethical discussions in analytics. However, as Willis stated in 2014, "…determining the most good for the most people is a responsible way forward, but perhaps not the most comprehensive, especially when assessing potential outcomes" (EDUCAUSE Review). A primary goal of analytics is to increase student success for the collective student body (current and future). This is a utilitarian rationale for the use of educational data. But there are other, less positive, potential outcomes of analytics.

Individuals, such as advisers or instructors, once informed of a student's predicted level of success, could inadvertently create circumstances that would cause the student to perform to the predicted level by, for example, unintentionally communicating the belief that the student is unlikely to succeed. Therefore, might a self-fulfilling prophesy be created when staff, faculty or students are made aware of predicted risk? Or when students are targeted for specific intervention(s) based on past performance or demographics associated with attrition? On the other hand, as Slade et al. pointed out, institutions have an obligation to "…act on knowledge gained through analytics". If there is evidence that, for example,

intervening with students predicted to be at risk for poor academic performance could significantly improve their learning, there is an ethical obligation to intervene.

This obligation is reinforced by federal regulations surrounding financial aid and employment after graduation that require institutions to report employment rates and wages for certificates/degrees for which students may use federal funds. The current state of analytics implies that the risk of creating a self-fulfilling prophesy is overshadowed by the possible individual/collective benefit of an ethical analytics implementation and the concurrent risk of failing to fulfill the institution's ethical responsibility to use available information to optimize students' chances for academic success.

Also assuaging fear of possible misuse of analytics is the recognition that advisors, instructors, and other institutional officials have always had access to data that could have been misused to assume the worst or create self-fulfilling prophecies regarding students. That is, higher education institutions made data such as prior coursework grades, High School GPAs, entrance exam and placement test scores available to university faculty/staff long before big data analytics existed. To mitigate risks of these data being misused institutions have long held policies and procedures regarding access to and use of student information.

Even so, carefully crafted messaging to staff, faculty and students is a must when employing analytics. How do we ensure students do not feel they have been reduced to a number or a predictive score or that circumstances beyond their control render them unlikely to graduate? How can we use the information we generate through analytics to create meaningful and productive conversations? How can we use this information to most efficiently guide students through their educational experience? Analytics must be employed with caution and in broad context in order to be effective.

In addition to analysis of student demographics, curriculum, learning, academic support programs, and other facets of the student experience, institutions should use data to continuously assess the effectiveness of policies, procedures, programs, services, and initiatives. There is an institutional responsibility, especially with scarce resources and escalating costs, to complete this type of introspective examination regularly. Only effective practices should be maintained. Further, as articulated at "The Asilomar Convention", institutions have a responsibility to share their evaluative outcomes to help shape the effectiveness of higher education overall. Students have many options for postsecondary education; continuous analysis of programs and services can be powerful in helping Reinvention Center member institutions to demonstrate the value, even in a digital age, of the education offered only at a residential research institution.

Distribution and Reporting of Analytics
While campuses have the capability of merging data from disparate sources and can analyze complex data sets, the optimal benefit is achieved only when the data and results are transformed into usable information. Information must be presented in meaningful ways in order to inform conversations and decision-making. This requires tailoring results for specific audiences and specific conversations. The days of producing one-size-fits-all retrospective reports each semester/year are over. Institutional researchers/analysts have a responsibility to provide data in far more timely and relevant ways than in the past. Information must be clear, concise, meaningful and, above all, accurate. It must inform about insights, trends, methodology and the practical significance of findings. Further, connections must be made between the results and institutional contexts such as budget or major initiatives.

In addition, individualized feedback to students and/or faculty should be consistently available. For example, a faculty member should have access to timely course analytics to better understand the impact of pedagogical adjustments or curricular changes. Students should be able to see whether changes in specific study habits were associated with increased academic success. Making these connections is a critical piece of successful analytics because they invite the campus community to be involved and to envision how analytics can support community members' individual goals.

These types of projects also allow research institutions to advance the scholarship of teaching, learning, and student success. Research institutions are in a unique position to do so. They can highlight the evidence-based practices that deepen student learning on their campus; they can help faculty to integrate their teaching and research responsibilities by encouraging research related to discipline-specific learning; and they can prioritize student learning by integrating the findings from such research into curricula, instructional delivery, and professional development. The Reinvention Center member institutions have a responsibility to be leaders in advancing these areas of scholarship.

Proposed Guidelines

Overall, the Reinvention Center member universities have many faculty, staff and administrators dedicated to promoting student success and deep learning through the informed use of analytics. Accompanying this dedication is an abundance of data, far more than at any point in the history of postsecondary education. In addition to the more traditional course demographics, expanded data on academic preparation and student success are also available within a given semester, rather than only after students have completed courses. Given this landscape, the following guidelines are proposed as a means to drive the meaningful and ethical use of data for analytics in higher education:

1. Institutions should establish, if there is not one already, a Data Governance committee with oversight of data collection, architecture, ethical use, and meaningful reporting. Membership should include IT and data authorities from each of the major functional areas on campus, as well as representatives from units and groups using such data to improve students' educational experiences, e.g., advising, student affairs, and academic support programs; curriculum and undergraduate affairs committees; and assessment and professional development offices. Institutional legal counsel should be involved or consulted regarding privacy and confidentiality issues.

2. Institutions should be fully transparent with students, faculty and staff about what types of data are collected, where/how it is stored and how the threat of a data breach is mitigated. This could be integrated with FERPA acknowledgements and training.

3. Institutions should use analytics responsibly in efforts related to the success of individual students, various subpopulations and the institution's broader mission. Responsible use includes carefully crafted, thorough training related to data and interpretation of results. When possible, data should be shared in aggregate form to minimize the sharing of files that contain personally identifiable information.

4. Institutions have an obligation to share results from research projects with their campus community to inform programs, services, and curriculum. They also have an obligation to share findings with the higher education community at large in order to further the scholarship related to teaching, learning, and student success. They should work with their IRBs to develop approaches that enable such contributions to this knowledge base while working within federal guidelines for the conduct of research involving human subjects.

Conclusion

The landscape of higher education is changing dramatically as the digital age advances. Certainly, data is more accessible. As institutions plan their analytics initiatives/projects there are many issues to consider. The current white paper highlights some of those issues and provides suggestions for how institutions might manage them. Above all, it is imperative that we move forward from a utilitarian and ethical framework that values students' rights and focuses on their potential to succeed in higher education.

References

The Asilomar Convention for Learning Research in Higher Education. (2015, October 19). Retrieved from http://asilomar-highered.info/

The Code of Federal Regulation (2009, January 15). Retrieved from http://www.hhs.gov/ohrp/humansubjects/guidance/45cfr46.html#46.101

Slade, Sharon and Prinsloo, Paul (2013). Learning analytics: ethical issues and dilemmas. American Behavioral Scientist, 57(10) pp. 1509–1528.

Willis, J. E. (2015, October 19). Learning Analytics and Ethics: A Framework beyond Utilitarianism [web document]. Retrieved from http://er.educause.edu/articles/2014/8/learning-analytics-and-ethics-a-framework-beyond-utilitarianism