Presented to the Interdisciplinary Studies Program:

UNIVERSITY OF OREGON
APPLIED INFORMATION MANAGEMENT

Applied Information Management
and the Graduate School of the
University of Oregon
in partial fulfillment of the
requirement for the degree of
Master of Science

Using Learning Analytics to Predict Academic Outcomes of First-year Students in Higher Education

CAPSTONE REPORT

Pete Sander
Manager, Information Services
Oregon State University

University of Oregon
Applied Information
Management
Program

Spring 2016

Academic Extension
1277 University of Oregon
Eugene, OR 97403-1277
(800) 824-2714

Approved by

Dr. Kara McFall
Director, AIM Program

Using Learning Analytics to Predict Academic Outcomes of First-year Students in Higher Education

Pete Sander

Oregon State University

Abstract

This annotated bibliography explores scholarly literature published between 2010 and 2016 that addresses the analysis of student-generated data, called learning analytics (Fiaidhi, 2014), with the intention of providing early intervention to promote better academic outcomes. It provides information to higher-education instructors and administrators who are interested in learning about (a) reducing attrition of first year students, (b) when the application of learning analytics produces the best results, and (c) predicting academic outcomes using learning analytics.

Keywords: big data, learning analytics, predict, higher education, LMS data

Table of Contents

Introduction to the Annotated Bibliography

Problem

Data generation occurs in almost every action taken by humans in the developed world. Data generation occurs continuously, whether it is clickstream data captured by an Internet browser session or GPS data generated by a mobile device while driving to the grocery store. This glut of data is creating the opportunity for organizations to use the data to inform decisions and tailor products and services to their clientele (Erevelles, Fukawa, & Swayne, 2016). People refer to this phenomenon as data analytics or “big data” (Clow, 2013). The field of higher education is also rife with the creation of, as well as the potential uses for, big data.

Similar to many aspects of modern life, generation of tremendous amounts of data occur in the everyday activities within higher education. Computer-based Learning Management Systems (LMSs) are ubiquitous within the field of higher education (Mijatovic, Cudanov, Jednak, & Kadijevich, 2013), which means that students generate data about their school-based activities. Estimates suggest that by the year 2020 there will be 32 billion devices connected to the internet generating data (Erevelles et al., 2016). Every mouse click, keystroke, and page view is recordable, creating an opportunity for collection and analysis of tremendous amounts of data. The analysis of this data can be extremely useful to educators and educational institutions because it offers opportunities for understanding how to optimize educational experiences that promote student success (Wagner & Ice, 2012).

Experts refer to the use of big data analysis to inform decisions in higher education as learning analytics (Fiaidhi, 2014). According to Clow (2013), the most common definition of learning analytics is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (p.685). In an effort to improve pedagogy, educators look to the data to help them understand what it takes to help produce a successful student. In the context of this study, a successful student is one who receives a C or better in the traditional assessment of the letter grading scale (A, B, C, D, and F).

According to the US Department of Education (2015), more than 40% of full-time first-time college students do not graduate within six years. Added to that, the cost of attending college has more than doubled over the last 30 years, even after taking inflation into account (U.S. Department of Education, 2015). While costs continue to increase, so does the importance of a postsecondary degree. By the year 2020, estimates suggest that two-thirds of jobs will require a postsecondary education (U.S. Department of Education, 2015). Graduation is a very important benchmark for a university since it means that tuition payments continue. Similarly, graduation is extremely important to the student, whose future earnings could hinge on their graduation. The U.S. Department of Education (2015) claims that over the course of a lifetime, students who earn a bachelor's degree will earn approximately one million dollars more than their peers without a postsecondary education. The student who drops out of a higher education institution does not gain the benefit of the diploma, but still has the burden of the cost of attending without earning a degree. In light of this, retention of students is very important to both the institution and the students themselves.

Using data analytics in higher education provides insight into a student's problem-solving skills, strategic learning choices, and other aspects of his or her performance (Thille et al., 2014). Since the student's academic performance is quantifiable, and learning analytics can provide the understanding of how the student arrived at that performance level, the use of this data can provide many conclusions (Thille et al., 2014). The use of learning analytics provides previously unavailable opportunities to evaluate learning theories, and learner feedback (Greller & Drachsler, 2012). It also allows for the establishment of early-warning systems, and development of future learning applications (Greller & Drachsler, 2012). Through learning analytics, institutions have the opportunity to improve pedagogy, thus also improving the likelihood of student retention.

While there are many potential uses of the data generated from learning analytics, this study focuses on the use of data analytics in higher education to improve student retention. This is accomplishable by predicting the future academic outcomes of students, as well as providing early intervention to help improve those predicted outcomes. The basis of prediction is potentially the student’s past work, his or her educational background, and/or historical data about statistically similar students who took the same classes, and/or enrolled in the same major of study. However, the prediction must occur early enough in the student’s career to allow for improvement within the current semester (or term). The relatively short window of time for improvement likely means that the analysis of data should occur in near real-time, or at least very shortly after the creation of data by the student. Specifically, this study explores whether data taken from a learning management system in real-time (or very near real-time) can be applied quickly enough to provide timely intervention of a struggling student to positively affect their academic outcome, thus improving the likelihood of retention of that student within the institution.

Purpose Statement

The purpose of this study is to explore the literature in order to examine whether big data analysis within the realm of higher education, often referred to as learning analytics (Fiaidhi, 2014), can be applied early enough in the students’ career to improve the likelihood of student retention. Applying learning analytics helps to (a) predict outcomes; (b) intervene when necessary; and therefore (c) improve the academic results of first year university students, thus leading to a higher retention rate of those students.

Research Question

Primary question. As the amount of data generated from students' activities within Learning Management Systems continues to grow, how can higher education institutions utilize these data to predict academic outcomes early enough in a term or semester to improve those outcomes of first year students?

Sub questions. When is the best time to intervene with a struggling student? What student activities tend to provide the most accurate predictions?

Audience

The intended audience for this report is primarily higher education instructors and higher education administrators who focus their attention on reducing the attrition of first year students. Data generated from learning analytics can provide educators with feedback on a student's current performance, and provide information to allow for making choices about the appropriate next actions (Thille et al., 2014), which can improve results for struggling students and therefore provide a higher likelihood of continued enrollment. By providing information on the timing and efficacy of the provision of this data, these educators can make informed decisions about the best time to intervene and specific areas of struggle. This research also provides potential benefit for academic counselors, students studying education and instructional design, curriculum specialists, higher education researchers, as well as education-related governmental agencies and for-profit education-related service providers (Greller & Drachsler, 2012).

Search Report

Search strategy. A search of the entire online UO Library (Search Scope: UO + Summit + Articles, etc.) is done to determine which results a full database approach provides. The specific searches use the Boolean logic *big data* AND *higher education*, followed by the Boolean logic *learning analytics* AND *higher education*. The search uses the following individual databases:

- ProQuest Social Sciences Premium Collection;
- ERIC (U.S. Dept. of Education);
- Science Citation Index Expanded (Web of Science);
- Social Sciences Citation Index (Web of Science);
- MEDLINE/PubMed (NLM);
- SAGE Journals;
- Sage Publications (CrossRef);
- Wiley Online Library;
- American Association for the Advancement of Science (CrossRef);
- Arts & Humanities Citation Index (Web of Science);
- Oxford Journals (Oxford University Press);
- CQ Researcher Online (CQ Press);
- PMC (PubMed Central);
- Informa – Taylor & Francis (CorssRef);
- SciVerse ScienceDirect (Elsevier); and
- IEEE Journals and Magazines.

The results of these two searches produces hundreds of articles that fit into the search criteria, but not necessarily into this specific topic. However, many of the seminal articles are present in this search result. The articles found in this search provide additional key words that prove useful in subsequent searches.

The next search queries the same Boolean logic, but instead done within the education-specific ERIC database, which is the database of journal and non-journal literature covering all areas in the field of education. The search uses the same Boolean logic as the first attempt (*big data* AND *higher education;* and *learning analytics* AND *higher education*), done in the ERIC database instead of as a query in the larger set of databases.

Similar queries allow finding additional resources. Through the same pattern of search techniques and with the same databases already noted, additional combinations of Boolean logic are used. Additional completed searches use the following terms, phrases, and words:

- big data;
- learning analytics;
- higher education;
- data mining;
- Learning Management System (as well as its acronym LMS);
- LMS data;
- analysis;
- predict;
- outcomes;
- retention;
- first year;

- early intervention; and
- real time.

When the results from the general database search become too broad due to a large number of results, the application of additional search filters will refine the search. The applied filters allow only peer-reviewed articles, only articles available in full-text, and only articles published between 2010 and 2016. The application of the filters creates a more manageable number of results to explore.

In addition to the online research, a personal interview with the person in charge of the learning analytics initiative at Oregon State University, Ms. Robin Pappas, was performed, but not electronically recorded. During said interview, Ms. Pappas provided several key resources of benchmark studies. Ms. Pappas provided known resources already explored within the scope of this study, as well as several resources not yet gathered. This meeting allowed for new perspectives on this problem, as well as new directions in which to take the research.

Information evaluation criteria. Scholarly sources are valuable when they adhere to five characteristics (University of Florida Center for Public Issues Education, 2014). Those characteristics are (a) authority, (b) timeliness, (c) quality, (d) relevancy, and (e) bias. In this study, peer-reviewed articles determine the authority of the author. Similarly, during the review of the article, notice was given to the list of references used by the author to ensure that the sources cited appeared to be academic in nature. In order to maintain a reasonable level of timeliness and currency, journals from the last five years take weight over those that are older. This appears to be a reasonable approach, and not difficult to adhere to, since big data analysis is a relatively new phenomenon. Noting the professional title of the author(s), and the organization that they represent, can control for relevancy and bias. This is determined by researching whether the source is scholarly; evidenced by where the author works, and whether the author received compensation by a potential beneficiary of the research.

Documentation approach. The approach taken to document the searches and results is two-fold, mainly because the initial method used was not as effective as possible. Initially, the quality results from the queries are hand-written in a notebook, noting the search terms, the databases searched, the author's name, and the title of the article. While this is effective in that it records information required, it is not as efficient as the second method adopted later in the process. Instead of handwriting those results, a Word document is used to record results. Added to the information collected in the first method is the digital object identifier (doi), as well as a copy of the author's abstract. The second approach described is much faster, and more thorough, therefore more likely an appropriate approach for the size of this project.

Annotated Bibliography

Introduction to annotated bibliography

The following annotated bibliography presents 16 references that discuss ideas on reducing the attrition of first year students in a university setting, the use of big data analytics to predict student outcomes, and whether the intervention can act as an early-warning system to detect poor-performing students early enough to improve outcomes. These references provide educators a framework by which to reduce student attrition by analyzing data generated from a student's activity in a learning management system; allowing for an intervention early enough in a term to make a difference in the student's academic outcome. There are three categories for presentation of the references designed to address specific aspects of this problem: (a) reducing attrition of first-year students, (b) determining at what point the application of learning analytics provides maximum efficacy, and (c) predicting outcomes using learning analytics. Each annotation includes (a) the complete bibliographic citation, (b) the published abstract by the author(s), and (c) a summary describing why the article is pertinent to the questions asked in this paper.

Category One: Reducing attrition of first-year students.

Haynes Stewart, T.L., Clifton, R.A., Daniels, L.M., Perry, R.P., Chipperfield, J.G., & Ruthig, J.C. (2011). Attributional retraining: Reducing the likelihood of failure. *Social Psychology of Education: An International Journal*, *14*(1), 75-92. doi: 10.1007/s11218-010-9130-2

Abstract. Failing a course is an acutely negative event for first-year university students, and a major contributor to high attrition rates at North American universities. Despite its

prevalence, course failure receives relatively little research attention. What can be done to reduce course failure and help first-year students remain in university? This study examined the efficacy of an Attributional Retraining treatment intervention to reduce course failure in an Introductory Psychology course. Attributional Retraining is designed to restructure students' causal explanations of poor performance by encouraging controllable attributions such as effort and strategy in place of immutable causes such as academic ability or intelligence. Relative to students in the control group, first-year students who received Attributional Retraining were less likely to fail the Introductory Psychology course (14.6% vs. 6.4%). This finding emerged beyond the effects of several well-established predictors of academic outcomes including student background characteristics (i.e., age, gender, and past academic performance) and learning environment variables (student registration status and participation in a first-year orientation program), suggesting the utility of Attributional Retraining for students with varying backgrounds and in different educational contexts. To the extent that Attributional Retraining is effective, inexpensive, and relatively easy to administer it may be a viable option for inclusion in orientation programs designed to reduce course failure and attrition among first-year university students.

Summary. This article focuses on whether Attributional Retraining is likely to improve outcomes in first-year students. Attributional Retraining (AR) is a theory that suggests that individuals are inclined to look for explanations for unexpected, negative, or important outcomes within their daily lives. Three general dimensions for classification of the explanations, or causal attributions, are: (a) locus of causality (whether the cause is internal or external to the individual); (b) stability (causes that are subject to change over time, or remain consistent), and; (c) controllability (causes that are controllable or not). Thus, causal attributions for academic outcomes directly affect motivation since they imply that performance is either controllable or uncontrollable. More simply stated, the more a student feels they have control over the situation, the more motivated that student will likely be to work diligently towards a successful outcome. Their specific study explores the efficacy of the use of AR interventions after controlling for the influence of background characteristics and learning environment variables. The study looks at 661 first-year students in a large Canadian university, of which 140 received AR interventions, and the rest (521) remain in the control group. Those students who receive AR interventions receive four separate treatments over the course of their first year. The study finds that students who received the AR interventions are 73% less likely to fail the course than those who did not receive the treatments. Since they find that the results emerge beyond the effects of several well-known predictors, an implication exists that suggests that the effects of AR may be significant across individual differences in background and learning environment variables. This study supports that attrition of first-year students is reducible through intervention, and therefore relates to the focus of this report.

Hudson, W.E. (2006) Can an early alert excessive absenteeism warning system be effective in retaining freshman students? *Journal of College Student Retention: Research, Theory, & Practice*, *7*(3-4), 217-226 doi: 10.2190/8TDY-798N-1ACK-9733

Abstract. A unilateral decision was made by the Office of the Provost and Executive Vice President for Academic Affairs, the Office of Academic Outreach and Support, and the Office of Academic Support and Retention (AS&R) to implement a pilot initiative for reporting, monitoring, and tracking excessive absenteeism during the Spring semester 2003. A Web-based system of notification was designed and provided to instructors for completion and submission. There were 216 students reported as having excessive absenteeism during the first six weeks of the spring semester. There were 78 instructors who submitted absenteeism reporting forms for 25 different courses. Ninety one students were successfully contacted by their advisors and responded. Of the students contacted and responded, 44 passed the course, 33 failed the course and 14 dropped the course. The implementation of this pilot study demonstrated its effectiveness by reducing the number of students who would have dropped, or failed courses due to a lack of attendance. Students reported that they were surprised to learn that their attendance was being monitored and amazed that someone cared enough to contact them and offer guidance and assistance. Faculty and staff were surprised that attendance was being monitored and that student and faculty contacts were initiated, responded to and reported back to the instructor.

Summary. This article discusses an early warning system at Morehead State University in Kentucky. The school implements a system whereby advisors contact students who were chronically absent in classes. The study finds that of the students contacted about their absenteeism, 48% of them were able to pass the class. The paper does not provide information about the 108 of the original 216 students who did not receive the early intervention, so it is difficult to know the efficacy of the program. Despite the lack of context around the efficacy, the author recommends four courses of action resulting from this study. The four recommendations are (a) enhance follow through and implement follow along procedures, (b) enhance the mechanisms for contacting and assisting advisors in locating students, (c) increase collaboration between departments, and (d) implement the early alert warning system on a continuous full time basis. This article provides examples on how to reduce attrition, and therefore helps to answer the questions posed in this study.

Lizzio, A., & Wilson, K. (2013). Early intervention to support the academic recovery of the first-year students at risk of non-continuation. *Innovations in Education and Teaching International*, *50*(2), 109-120. http://dx.doi.org/10.1080/14703297.2012.760867

Abstract. The widening participation agenda and related concerns about student retention require a more systematic focus on supporting student success. This paper describes a process designed with the dual goals of supporting the short-term academic recovery of students at risk of non-continuation due to early difficulties with assessment and developing their ongoing capabilities for self-regulation around challenging assessment tasks. Commencing students who failed or marginally passed their first piece of university assessment were invited to participate in a two-stage process: independently completing a reflective workbook designed to help them understand the reasons for their assessment performance, followed by a structured consultation with their tutor to identify improvement goals and strategies. Students undertaking the academic recovery process achieved higher pass rates for the second assessment item and for the course overall than a comparative group of students who did not participate. Findings indicate that a self-regulation-based intervention can contribute to students' academic persistence and success. Importantly, students experienced the intervention as providing insight into their underperformance on assessment and developing their capacity for metalearning.

Summary. In this paper, the authors discuss a study performed at Griffith University in Australia. The study examines 250 students enrolled in an introductory psychology course over two separate cohorts over a two-year period. An intervention takes place after the first exam during week six of the semester, on students who receive a failing grade (below 50%) or a just achieve a passing grade (50%-55%). The intervention occurs the following week, informing students that the study was purely voluntary, and their lack of participation will not affect their grade. Overall, those who participate in the consultation benefit from the action, as 63% of cohort 1 and 60% of cohort 2 pass the course, whereas only 26% of cohort 1 and 24% of cohort 2 who did not participate in the consultation pass the course. The research suggests that early intervention of struggling students does appear to provide a positive influence on the outcomes of students.

Singell, L.D., & Waddell, G.R. (2010). Modeling retention at a large public university: Can at-risk students be identified early enough to treat? *Research in Higher Education, 51*(6), 546-572. http://dx.doi.org/10.1007/s11162-010-9170-7

Abstract. We examine the extent to which readily available data at a large public university can be used to a priori identify at-risk students who may benefit from targeted retention efforts. Although it is possible to identify such students, there remains an inevitable tradeoff in any resource allocation between not treating the students who are likely to exit without treatment and treating students who are likely not to exit in the absence of the treatment. At-risk students are found to remain at risk throughout their college career. Moreover, conditional on exiting the institution, the degree to which the student was at risk is predictive of whether the student subsequently re-enrolls elsewhere and the type of institution at which this re-enrollment occurs. In this context, we discuss how retention policies relate to insuring the initial match is appropriate, recognizing that some attrition can be in keeping with the broad social interest.

Summary. This article explores the retention rates of first year students at the University of Oregon, sharing an evaluation of cohorts of incoming fall term freshman between 2001 and 2006. Using common attributes (e.g. high school GPA, race, gender, whether the student receives financial aid, etc.) to group the students, the researchers classify students into 10 deciles based on initial enrollment statistics. The researchers create the deciles by placing the students in a ranked order based on likelihood of attrition, then group them based on that ranking into 10 equal parts. They follow each decile through their career at the University, and researchers note their activity regarding enrollment. They know that historically the University graduates slightly more than 60% of students who enroll as freshmen by their sixth year. The statistics about the 2001 cohort agree with that historical data in that 1780 of 2851 (62.4%) graduated as of the winter term of 2008. They also find that of those who started at the University of Oregon, roughly 89% of students who graduated did not take time off through their fifth year of enrollment. Perhaps most importantly, they find that 39% of the attrition from the 2001 cohort occurs before the fall term of their sophomore year. They conclude that keeping students actively enrolled past the first term is critical to their overall graduation rates. The authors suggest that their model tends to identify students who are most vulnerable to not being retained, providing the opportunity to intervene. Specifically, the authors suggest that at-risk students are identifiable by analyzing information available at the time of student’s enrollment. However, while this information is helpful, the authors suggest that a controlled experiment should be performed to determine whether interventions (treatments) to a subset of at-risk students provides a significant benefit related to those at-risk students who were not randomly selected for intervention. This paper suggests that at-risk students are indeed identifiable, and that attrition is most prevalent in first-year students, and therefore a good target for intervention. As a result, this article provides evidence to further the study within this paper.

Vander Schee, B.A. (2011). Early intervention: Using assessment to reduce student attrition. *About Campus, 16*(1), 24-26. http://dx.doi.org/10.1002/abc.20051

Abstract. The number of studies conducted on college-student attrition is overwhelming. But few examine the impact of adding an early-intervention assessment tool to existing retention programs. Too often, colleges and universities conduct initiatives with similar purposes as disconnected efforts; retention programs in particular can benefit from a more coordinated approach. In addition, bringing existing efforts together can produce significant results with minimal resistance to change. This article presents a case study of an institution that combined two previously unrelated retention initiatives in a way that dramatically increased the effectiveness of both.

Summary. This article is a case study from a small liberal arts college in the Northeast United States that implements an early intervention program for incoming freshman at the college. The school uses a College Student Inventory (CSI) survey, as well as offering an elective three-credit freshman seminar to first-year students. As part of an initiative to decrease attrition, the college implements a pilot program that joins the CSI and freshman seminar for some freshman, but not for others. The result is that those who enroll in the pilot that includes the CSI has a first-to-second year retention of 84.6%, compared to 67.9% of those who did not take the course. Based on those results, the college requires the CSI and freshman seminar as requirements the following year, which increases the overall first- to second-year retention from 68.6% to 71.2%. These results show that early intervention of students does indeed lead to improved retention, and therefore helps support the answer to the questions asked in this study.

Category Two: At what point should learning analytics be applied for maximum efficacy?

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society, 15*(3), 42-57. Retrieved from http://www.jstor.org/stable/jeductechsoci.15.3.42

Abstract. With the increase in available educational data, it is expected that Learning Analytics will become a powerful means to inform and support learners, teachers and their institutions in better understanding and predicting personal learning needs and performance. However, the processes and requirements behind the beneficial application of Learning and Knowledge Analytics as well as the consequences for learning and teaching are still far from being understood. In this paper, we explore the key dimensions of Learning Analytics (LA), the critical problem zones, and some potential dangers to the beneficial exploitation of educational data. We propose and discuss a generic design framework that can act as a useful guide for setting up Learning Analytics services in support of educational practice and learner guidance, in quality assurance, curriculum development, and in improving teacher effectiveness and efficiency. Furthermore, the presented article intends to inform about soft barriers and limitations of Learning Analytics. We identify the required skills and competences that make meaningful use of Learning Analytics data possible to overcome gaps in interpretation literacy among educational stakeholders. We also discuss privacy and ethical issues and suggest ways in which these issues can be addressed through policy guidelines and best practice examples.

Summary. This article presents a framework upon which a learning analytics platform is most effective. The framework builds on six dimensions that require consideration in every case of establishing a learning analytics initiative. Those six dimensions are (a) stakeholders, (b) objectives, (c) data, (d) instruments, (e) external constraints, and (f) internal limitations. The application of learning analytics potentially leads to an enhanced learning process and personalized information about students by uncovering and analyzing previously hidden information out of educational data. The desired result is a system that allows for predicting and modeling learner activity, which can lead to earlier intervention or to adaptive services and curricula. There are, however, challenges and ethical concerns for consideration. The authors warn that data analyzers cannot allow established biases and prejudices (for example age, race, socioeconomic status) to cloud their judgment. Furthermore, the same set of data is interpretable in many ways, potentially providing multiple theories, which could lead to varying (and perhaps conflicting) decisions on how to use the data by an instructor or advisor. This article directly relates to the question posed in this study, describing how and when to use learning analytics.

Macfadyen, L.P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education, 54*(2), pp.588-599. http://dx.doi.org/10.1016/j.compedu.2009.09.008

Abstract. Earlier studies have suggested that higher education institutions could harness the predictive power of Learning Management System (LMS) data to develop reporting tools that identify at-risk students and allow for more timely pedagogical interventions. This paper confirms and extends this proposition by providing data from an international research project investigating which student online activities accurately predict academic achievement. Analysis of LMS tracking data from a Blackboard Vista-supported course identified 15 variables demonstrating a significant simple correlation with student final grade. Regression modelling generated a best-fit predictive model for this course which incorporates key variables such as *total number of discussion messages posted*, *total number of mail messages sent*, and *total number of assessments completed* and which explains more than 30% of the variation in student final grade. Logistic modelling demonstrated the predictive power of this model, which correctly identified 81% of students who achieved a failing grade. Moreover, network analysis of course discussion forums afforded insight into the development of the student learning community by identifying disconnected students, patterns of student-to-student communication, and instructor positioning within the network. This study affirms that pedagogically meaningful information can be extracted from LMS-generated student tracking data, and discusses how these findings are informing the development of a customizable dashboard-like reporting tool for educators that will extract and visualize real-time data on student engagement and likelihood of success.

Summary. This article presents a proof of concept and the foundational blocks upon which to build an early warning system, allowing an intervention of students at risk of a poor outcome in a course. The specific foundations of an early warning system are (a) Internet and communication technologies integrated into teaching and learning, (b) increased ability of LMS tracking data, (c) the emergence of academic analytics, and (d)

increased attention to the social nature of learning. The study examines 118 students over three terms (five total classes) within a fully online undergraduate biology course at the University of British Columbia. The study demonstrates a positive correlation between several LMS data variables and the final grade of the student, with the best predictors being (a) total number of discussion messages posted, (b) total number of mail messages sent, and (c) total number of assessments completed. The data suggest that a tool that extracts and visualizes real-time pedagogically meaningful information on student engagement and the likelihood of success would be invaluable for contemporary educators.

Tempelaar, D.T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167. http://dx.doi.org/10.1016/j.chb.2014.05.038

Abstract. Learning analytics seek to enhance the learning processes through systematic measurements of learning related data and to provide informative feedback to learners and teachers. Track data from learning management systems (LMS) constitute a main data source for learning analytics. This empirical contribution provides an application of Buckingham Shum and Deakin Crick's theoretical framework of dispositional learning analytics: an infrastructure that combines learning dispositions data with data extracted from computer-assisted, formative assessments and LMSs. In a large introductory quantitative methods module, 922 students were enrolled in a module based on the principles of blended learning, combining face-to-face problem-based learning sessions with e-tutorials. We investigated the predictive power of learning dispositions, outcomes of continuous formative assessments and other system generated data in modelling

student performance of and their potential to generate informative feedback. Using a dynamic, longitudinal perspective, computer-assisted formative assessments seem to be the best predictor for detecting underperforming students and academic performance, while basic LMS data did not substantially predict learning. If timely feedback is crucial, both use-intensity related track data from e-tutorial systems, and learning dispositions, are valuable sources for feedback generation.

Summary. This article describes an empirical study that explores the value of data sources with regard to their predictive capabilities and timeliness. Instead of only exploring the data generated by student activity within a LMS, this study takes into consideration the use of e-tutorials and other learning formats of blended learning and how they affect performance, as well as how feedback based on learning dispositions affect learning outcomes. The study uses data from Blackboard (a popular LMS), MyLabs (an e-tutorial environment for test-directed learning and practicing), learning disposition data (learning styles, learning motivation, and learning emotions), demographic factors of the students, entry test information, and academic performance; overall exploring 102 different variables. They find that tracking data from the LMS alone is not a strong indicator of under-performing students. The best predictor, based on this study, is data generated by the e-tutorial assessments; however, the generation of this type of data can tend to be late in the term, therefore reducing the efficacy of the predictive power simply because there is not sufficient time to intervene with a struggling student before the term ends. The authors suggest that prior to the middle of the term (in the case of their study, week three of eight) is the most effective time to intervene with a struggling student. This is because enough data are present to see a clear picture of the student's work, but there is still enough time to improve outcomes. The authors conclude that the best mix of data comes from a combination of e-tutorial (quizzes and practice tests) and learning dispositions if timely feedback to the student is required. The results of this study support the claim that feedback must be timely and accurate to be effective. Caveat Emptor: Funding for this study is partially from the Dutch SURF-foundation, which is part of the Learning Analytics Stimulus program. This could result in pressure to produce results that suggest a greater value in learning analytics than what actually occur.

Thille, C., Schneider, E., Kizilcec, R., Piech, C., Halawa, S., & Greene, D.K. (2014). The future of data-enriched assessment. *Research & Practice in Assessment, 9*. 5-16. http://www.rpajournal.com/the-future-of-data-enriched-assessment/

Abstract. The article addresses the question of how the assessment process with large-scale data derived from online learning environments will be different from the assessment process without it. Following an explanation of big data and how it is different from previously available learner data, we describe three notable features that characterize assessment with big data and provide three case studies that exemplify the potential of these features. The three case studies are set in different kinds of online learning environments: an online environment with interactive exercises and intelligent tutoring, an online programming practice environment, and a massive open online course (MOOC). Every interaction in online environments can be recorded and, thereby, offer an unprecedented amount of data about the processes of learning. We argue that big data enriches the assessment process by enabling the continuous diagnosis of learners' knowledge and related states, and by promoting learning through targeted feedback.

Summary. This article includes three case studies of the use of learning analytics at Stanford University, and ways big data helps assess learning activities and promotes student success. There are three elements of data-enriched assessment, which are (a) the ability to observe student activity continuously, (b) the ability to provide feedback to learners about current activities and appropriate next actions, and (c) the ability to explore multiple facets of the learning process by exploring data generated by students. The case studies each address a different benefit of learning analytics. The first case study explores the open learning initiative, and gathers data from over 1,000,000 learners. The study looks at ways to assess skill in real-time by using machine-learning algorithms. The second case study analyzes logs of open-ended programming assignments with the intention to (a) explore new perspectives into individual learner abilities, (b) show how learners approach a problem, and (c) understand how learners navigate complex assignments. In this study, the authors assess how learners worked through the problem, as opposed to just the final answer to the problem. The third case study explores predicting a student dropping out based on their interaction with the learning management system. Specifically, they analyze the amount of time spent to complete the first week's videos, the scores from the assignments, and the fraction of videos and assignments skipped. They were able to predict dropouts at a 93% recall rate. They believe that big data can enrich assessment for three reasons. First, online learning environments allow for the continuous collection of data at a large scale. Second, it allows educators to record multifaceted measurements of skills and tendencies that normally evade traditional assessments. Finally, online learning environments are capable of delivering personalized feedback at the right moment. This article supports the use of learning analytics to provide real-time feedback to students.

Category Three: Predicting outcomes using learning analytics.

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. http://www.ncolr.org/jiol/issues/pdf/12.1.2.pdf

Abstract. Learning analytics is receiving increased attention, in part because it offers to assist educational institutions in increasing student retention, improving student success, and easing the burden of accountability. Although these large-scale issues are worthy of consideration, faculty might also be interested in how they can use learning analytics in their own courses to help their students succeed. In this paper, we define learning analytics, how it has been used in educational institutions, what learning analytics tools are available, and how faculty can make use of data in their courses to monitor and predict student performance. Finally, we discuss several issues and concerns with the use of learning analytics in higher education.

Summary. This article explores the research and provides suggestions on how faculty can utilize learning analytics in order to improve student outcomes. The authors discuss past research, provide suggestions on usage, and describe why learning analytics are useful to administrators, instructors, and students alike. The benefit to the institution is an increased level of accountability, the improvement of decision-making and informing resource allocation, highlighting institutional successes and challenges, as well as reducing attrition of students. The benefit to instructors is the allowance of data-driven prediction of learner performance, testing and evaluating curricula, and developing effective instructional techniques. For the student, the benefits are improving academic success, detecting undesirable learning behaviors, and allowing an insight into their own learning styles and behaviors. While there are many benefits to learning analytics, institutions and instructors need to have an understanding of the shortcomings and potential issues that can arise. A primary point is that pedagogy should drive learning analytics, not the converse. In addition, students can feel as though they are being monitored, which can be threatening to some. The analysis of LMS data can be helpful, but it is not a holistic view, as it does not take into account issues like interpersonal problems. Data privacy is also a concern, and needs careful consideration before a learning analytics initiative is undertaken. The final shortcoming presented is that institutions need to self-reflect on whether the data is used to measure learning, or whether it is a thinly veiled attempt to boost student retention and course completion. The authors conclude that learning analytics are here to stay, and therefore institutions and instructors need to understand the benefits and drawbacks associated. This article addresses learning analytics from the faculty perspective, discussing the benefits and drawbacks, but generally does support the claim that learning analytics can predict student outcomes.

Hu, Y., Lo, C., & Shih, S. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469-478. http://dx.doi.org/10.1016/j.chb.2014.04.002

Abstract. An early warning system can help to identify at-risk students, or predict student learning performance by analyzing learning portfolios recorded in a learning management system (LMS). Although previous studies have shown the applicability of determining learner behaviors from an LMS, most investigated datasets are not assembled from online learning courses or from whole learning activities undertaken on courses that can be analyzed to evaluate students' academic achievement. Previous studies generally focus on the construction of predictors for learner performance evaluation after a course has ended, and neglect the practical value of an "early warning" system to predict at-risk students while a course is in progress. We collected the complete learning activities of an online undergraduate course and applied data-mining techniques to develop an early warning system. Our results showed that, time-dependent variables extracted from LMS are critical factors for online learning. After students have used an LMS for a period of time, our early warning system effectively characterizes their current learning performance. Data-mining techniques are useful in the construction of early warning systems; based on our experimental results, classification and regression tree (CART), supplemented by AdaBoost is the best classifier for the evaluation of learning performance investigated by this study.

Summary. This article describes a framework for the development of an early warning system including time-dependent variables that can predict student performance through data mining from an LMS. By using data collected on 14 variables from an LMS containing 300 undergraduate students taking the Information Literacy and Information Ethics online course in a national university in Taiwan, the study explores student behaviors at three separate times during the course (weeks four, eight, and thirteen) to determine how the early warning system can accurately predict learning performance. The authors categorized the specific types of input variables as (a) login behavior, (b) the use of online course materials, (c) assignment status, and (d) discussion status in the forum. Their results conclude that regardless of classification techniques, the accuracy of an early warning system is improvable by considering time-dependent variables. An added benefit to this system is that instructors can use the data to improve their teaching methods in a timely fashion, and learners can compare their learning history to their peers within the University. This comparison allows for a self-monitoring of learning performance. The study helps corroborate the claim that the timeliness of feedback is critical to determining students who are struggling in order to improve academic outcomes.

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *Internet & Higher Education*, *27*. 54-63. http://dx.doi.org/10.1016/j.iheduc.2015.06.001

Abstract. Digital textbook analytics are a new method of collecting student-generated data in order to build predictive models of student success. Previous research using self-report or laboratory measures of reading show that engagement with the textbook was related to student learning outcomes. We hypothesized that an engagement index based on digital textbook usage data would predict student course grades. Linear regression analyses were conducted using data from 233 students to determine whether digital textbook usage metrics predicted final course grades. A calculated linear index of textbook usage metrics was significantly predictive of final course grades and was a stronger predictor of course outcomes than previous academic achievement. However, time spent reading, one of the variables that make up the index was more strongly predictive of course outcomes. Additionally, students who were in the top 10th percentile in number of highlights had significantly higher course grades than those in the lower 90th percentile. These findings suggest that digital textbook analytics are an effective early warning system to identify students at risk of academic failure. These data can be collected unobtrusively and automatically and provide stronger prediction of outcomes than prior academic achievement (which to this point has been the single strongest predictor of student success).

Summary. This article discusses the benefits of using digital textbook usage data as a predictor of student outcomes. Digital textbooks can provide data on student engagement through the inclusion of embedded quizzes, the capturing of page views, time spent reading, number of sessions a student engages with the text, as well as highlighting and note-taking activities. Research explains that students who engage more in the class activities, like reading the assigned text, perform better in the course. Understanding this, a student's engagement with a digital textbook can be a worthwhile predictor. The authors suggest that because the student does not realize that they are being monitored (data collected on their usage of, and activities within the textbook), the data generated is highly authentic and reliable, and therefore can act as a proxy for a student's reading skills and learning activities. Their study examines 236 students from Texas A&M University – San Antonio, who received digital textbooks for this study. The research shows that engagement with the digital textbook was a stronger predictor of course grades than a student's previous grades. Experts generally agree that previous grades are also a very reliable predictor. Another benefit of digital textbook engagement analytics is that it can occur in real time, providing an opportunity to engage with a student before it is too late. The authors suggest that digital textbook analytics can be useful as an effective early warning system to identify students in danger of a poor academic outcome. This study supports the idea that the use of learning analytics can act as a useful predictor of student academic outcomes.

Niemi, D., & Gitin, E. (2012). Using big data to predict student dropouts: Technology affordances for research. *International Association for Development of the Information Society, Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in Digital Age (CELDA) (Madrid, Spain, Oct 19-21, 2012). 4 pp.*

Abstract. An underlying theme of this paper is that it can be easier and more efficient to conduct valid and effective research studies in online environments than in traditional classrooms. Taking advantage of the "big data" available in an online university, we conducted a study in which a massive online database was used to predict student successes and failures. We found that a pattern of declining performance over time is a good predictor of the likelihood of dropping out, and that having dependents or being married or in the military reduces the risk of dropping out. The risk of dropping out was higher for older students, females, and students with previous college education or transfer credits. These results provide a foundation for testing interventions to help students who are at risk and will also help to inform the development of a "research pipeline" that will enable rapid experimental studies of new tools and strategies.

Summary. In this study, the authors explore an existing database of 14,791 students enrolled in a fully online program. They explore how both student demographics (e.g. age, gender, race, marital status, military status, previous college education, estimated family financial contribution, and the number of transfer credits from another university), as well as academic variables (e.g. final exam score, discussion participation, project scores, and other assignment scores) relate to dropout rates. The study reveals that the decline in student performance is a significant predictor of the likelihood of dropping out. Analysis of the student demographics and academic variables provides a percentage of increased or decreased likelihood of dropping out. The authors are surprised by what the data suggested about certain populations and their likelihood to drop out; specifically, (a) females are at a 330% increased risk of dropping out, (b) married students are at a 35.7% reduced risk of dropping out, and, (c) students with transfer credits are at a 236% increased risk of dropping out. They conclude that further research is needed, and furthermore, testing of new instructional, motivational and support strategies to help struggling students is needed to improve learning analytics and online education in general. This article supports the use of data to predict academic outcomes of students.

Smith, V.C., Lange, A., & Huston, D.R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, *16*(3), 51-61. **http://sloanconsortium.org/jaln/v16n3/predictive-modeling-forecast-student-outcomes-and-drive-effective-interventions-online-co**

Abstract. Community colleges continue to experience growth in online courses. This growth reflects the need to increase the numbers of students who complete certificates or degrees. Retaining online students, not to mention assuring their success, is a challenge that must be addressed through practical institutional responses. By leveraging existing student information, higher education institutions can build statistical models, or learning analytics, to forecast student outcomes. This is a case study from a community college utilizing learning analytics and the development of predictive models to identify at-risk students based on dozens of key variables.

Summary. This article is a case study of the use of learning analytics at Rio Salado Community College with the intended result of predicting student outcomes in order to establish an early-warning system for poor-performing students. The authors focus their study on just one online freshman accounting course, which provides a sample of 539 students made from only online students who receive a final grade for the course. The data extracted from the LMS for analysis of the generated student activity and interaction with the LMS. Examples of those activities are logging in to the LMS, viewing a grade, completing an assessment, and opening a lesson. This study also weighed recent activity more heavily than activities from earlier weeks of the course. The output of the analysis is a three-level warning system (low, medium, high), providing instructors a user-friendly abstraction of the estimated probability of course success. The addition of weighting resulted in an increased accuracy in predicting course outcomes as information accumulates over the duration of the course. The authors also discuss the implementation of the intervention processes, which were (a) a phone call from the instructor to the struggling student, and (b) automated course welcome emails to students the day before a course started. While these provide evidence of improved results, the data show ineffectiveness or unrepeatable results. The authors conclude that a strong correlation exists between LMS activity and course outcomes. This supports the theory that learning analytics can predict academic outcomes of students. Caveat Emptor: One of the authors of this study worked for Ellucian at the time of publishing. Ellucian is a for-profit company that provides technology services to higher-education institutions to achieve

student success. This should not affect the results of this paper, since it appears that at the time of the data collection he was a reporting analyst/programmer in Institutional Research at Rio Salado College.

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior*, *47*, 168-181. http://dx.doi.org/10.1016/j.chb.2014.09.034

Abstract. Building a student performance prediction model that is both practical and understandable for users is a challenging task fraught with confounding factors to collect and measure. Most current prediction models are difficult for teachers to interpret. This poses significant problems for model use (e.g. personalizing education and intervention) as well as model evaluation. In this paper, we synthesize learning analytics approaches, educational data mining (EDM) and HCI theory to explore the development of more usable prediction models and prediction model representations using data from a collaborative geometry problem solving environment: Virtual Math Teams with Geogebra (VMTwG). First, based on theory proposed by Hrastinski (2009) establishing online learning as online participation, we operationalized activity theory to holistically quantify students' participation in the CSCL (Computer-supported Collaborative Learning) course. As a result, 6 variables, *Subject*, *Rules*, *Tools*, *Division of Labor*, *Community*, and *Object*, are constructed. This analysis of variables prior to the application of a model distinguishes our approach from prior approaches (feature selection, Ad-hoc guesswork etc.). The approach described diminishes data dimensionality and systematically contextualizes data in a semantic background.

Secondly, an advanced modeling technique, Genetic Programming (GP), underlies the developed prediction model. We demonstrate how connecting the structure of VMTwG trace data to a theoretical framework and processing that data using the GP algorithmic approach outperforms traditional models in prediction rate and interpretability. Theoretical and practical implications are then discussed.

Summary. In this article, the authors describe a prediction model framework that they suggest is easier for instructors to use, but still statistically robust enough to be able to provide meaningful results. They suggest that most statistics-based prediction models in use now are difficult to interpret, which is problematic for two reasons. First, a “black box” (a complex piece of equipment with unknown contents to the user) make predictions without explanation, which can lead to a lack of confidence by the instructor. Second, if a model is not understandable, users may not be able to validate it. The framework the authors provide emphasizes participation in online learning as a central factor affecting performance. They study electronic trace data from an online math course (virtual math team, or VMT) with 122 students, which took place in 2012-2013. They use cluster analysis applied to the student’s learning outcomes to generate categories for student performance. While they suggest there are other models that produce a slightly more accurate prediction, the ease-of-use factor allows for a wider adoption by instructors who may not have a background in data analytics. In addition, the analysis allows for a more granular level of examination, providing an opportunity to provide more concrete and individualized suggestions to each student. The authors point out some limitations to their research. First, the model does not take into consideration qualitative aspects of collaboration by the students, and secondly, researchers with less familiarity with the VMT environment may have a difficult time replicating the results in a different context. This study suggests that prediction of student outcomes by using data analytics is becoming increasingly easier to do, without negatively affecting accuracy.

Zacharis, N.Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *Internet and Higher Education*, *27*, 44-53. http://dx.doi.org/10.1016/j.iheduc.2015.05.002

Abstract. This study aimed to develop a practical model for predicting students at risk of performing poorly in blended learning courses. Previous research suggests that analyzing usage data stored in the log files of modern Learning Management Systems (LMSs) would allow teachers to develop timely, evidence-based interventions to support at risk or struggling students. The analysis of students' tracking data from a Moodle LMS-supported blended learning course was the focus of this research in an effort to identify significant correlations between different online activities and course grade. Out of 29 LMS usage variables, 14 were found to be significant and were input in a stepwise multivariate regression which revealed that only four variables – Reading and posting messages, Content creation contribution, Quiz efforts and Number of files viewed – predicted 52% of the variance in the final student grade.

Summary. The focus of this study is analyzing multiple variables, as opposed to just one, to determine students at risk of earning a failing grade. The author studies 134 freshmen students studying Computer Science and Computer Engineering who take an introductory Computer Science course taught in a hybrid environment; meaning that it is taught partially online and partially in a more traditional face-to-face environment. The author finds that there are 14 activities/variables showing significant relationship with the final grade of the student, but four variables appear as the best predictors. The four most significant variables that the author determine as the most predictive in student outcomes are (a) reading and posting messages on the discussion board, (b) content creation contribution, (c) quiz engagement, and (d) files viewed online. The results show that these variables provide an 81.3% accuracy in prediction. This study varies from past studies in that it uses a more complex analysis method combining multiple variables as opposed to just a single variable. The results of this study provide corroboration that the analysis of student data from an LMS can provide significant indication of success of a student.

Conclusion

While the analysis of big data in an education setting, often referred to as learning analytics (Fiaidhi, 2014), is a relatively new phenomenon, it is not something that is likely to go away anytime soon. In fact, with the cost to the student increasing in higher education (U.S. Department of Education, 2015), and the increased pressure on higher educational institutions to control costs and reduce attrition (Fiaidhi, 2014), it is likely that learning analytics will increase in importance.

Learning analytics provide an institution with insight into a learner's problem-solving skills and other aspects of the student's learning profile (Thille et al., 2014), while also providing the institution with an opportunity to improve teaching and overall pedagogical approach (Hu, Lo, & Shih, 2014). With this in mind, there are many potential benefits from using learning analytics in higher education. One of those benefits is providing an early-warning system to discover poor-performing students with the intention of providing an intervention in hope of retaining the student within the institution (Greller & Drachsler, 2012; Macfadyen & Dawson, 2009). In order to understand how and why an early intervention is possible, the research explores three general topics in current literature: (a) reducing attrition of first-year students; (b) applying learning analytics for maximum efficacy, and; (c) predicting outcomes using learning analytics.

Reducing attrition of first-year students

The retention of first-year students is an issue that higher education should focus on in order to maintain a higher overall enrollment, and therefore retain higher tuition revenue. Studies show that the bulk of attrition occurs between the freshman and sophomore years at an institution (Singell & Waddell, 2010; Vander Schee, 2011). Specifically, Singell and Waddell (2010) found that 39% of attrition occurs before the fall term of the student's sophomore year, concluding that keeping a student past their first year is critical to their long-term success, and ultimately graduation from the institution. In order for institutions to maintain retention levels, students must be successful enough to pass their classes, and reduce their academic failures. Reducing small-scale academic failures will ultimately lead to a lower attrition of students (Haynes et al., 2011).

In order to reduce small-scale failures of first-year students, researchers make several suggestions to produce positive results. Vander Schee (2011) suggests that the use of a freshman seminar class (often also referred to as a freshman orientation) can improve first-year results, and found that when the school required first-year students to take freshman seminar, they saw an increased retention of 2.6% (from 68.6% to 71.2%) in the first year of the pilot program. Similarly, Haynes et al. (2011) found that students are 73% less likely to fail a course if they receive routine interventions from institutional representatives. Additionally, the intervention of poorly performing students produces a clearly positive influence on the overall student success in Lizzio and Wilson's study (2013), citing that the students who receive the intervention are much more likely to pass the course.

In order to improve student success, and therefore reduce attrition, Hudson (2006) recommends that institutions (a) enhance follow through and implement follow along procedures, (b) enhance the mechanisms for contacting and assisting advisors in locating students, (c) increase collaboration between departments, and (d) implement the early alert warning system on a continuous full time basis. The use of an early-intervention program positively influences academic outcomes by first-year students (Lizzio & Wilson, 2013), and therefore is a good method by which an institution can reduce the attrition of first year students.

At what point should learning analytics be applied for maximum efficacy

The use of learning analytics to predict outcomes is potentially something that can benefit students, instructors, and institutions. However, the timeliness of when the analysis is completed, and/or the prediction made, is extremely important to the efficacy of the actions taken. The finest analysis of the most accurate data is not terribly helpful to a student if the information reaches the student after the completion (perhaps failure) of the class. Tempelaar et al. (2015) posit that the power of predictions made early in a term are about as accurate as predictions made later in the class session. Since the benefit of waiting to apply additional analytics is not significantly higher, interventions made early are much better than those made later in the class. Specifically, Tempelaar et al. (2015) suggests that week three, in an eight-week class, is the best time to intervene with a student.

The end goal of the use of learning analytics is to allow monitoring of student engagement in real-time (Macfadyen & Dawson, 2010), which means that the learning process is continuously observable (Thille et al., 2015). The intention behind mining student data is that it is useful at any time during a course (Macfadyen & Dawson, 2010), and that the use of learning analytics leads to an early intervention of students (Greller & Drachsler, 2012).

Learning analytics support decision-making (Greller & Drachsler, 2012). The basis of making decision about students is on instructional design, and overall institutional direction (Greller & Drachsler, 2012; Thille et al, 2015). What is potentially so powerful about the use of learning analytics is that it provides instructors an opportunity to see *how* a student came up with an answer, not just the answer itself (Thille et al., 2014). The ability to examine the process by which a student derives an answer to a problem is in itself predictive of their learning competencies (Thille et al., 2015). When applied in a timely fashion, learning analytics are able to predict potential student outcomes quickly enough to improve the outcomes of that student (Macfadyen & Dawson, 2010).

Predicting outcomes using learning analytics

One of the benefits of analyzing learning data is the potential ability to predict academic outcomes of students. In the past, before learning analytics were possible, it forced instructors to rely on their own intuition and hunches to determine which of their students are struggling (Dietz-Uhler & Hurn, 2013). Certainly, a higher degree of student-teacher interaction will typically lead to more student engagement, motivation, and achievement (Zacharis, 2015). Through learning analytics, an instructor has the opportunity to become more familiar with the students' learning habits, and therefore will have the opportunity to intervene as problems arise, potentially improving the success of students. As student success increases, retention of students also tends to increase, and as a result, the accountability of the university improves as well (Dietz-Uhler & Hurn, 2013).

Through the examination of student-produced data, for example, time spent logged into an LMS, quizzes taken, and content created, instructors are able to analyze whether a student is likely to succeed or fail (Zacharis, 2015). Similarly, digital textbook usage tracks student engagement at any given time during a course, and is useful in predicting the success of a student (Junco & Clem, 2015). The accuracy of the prediction will depend on the quality of the data, as well as the quality of the analysis. However, Hu, Lo, and Shih (2014) found that including time-dependent variables increase the accuracy of the early warning system.

Students are not the only benefactor of learning analytics and predictions based on data. Instructors can use LMS and other data to create a statistical model of a successful student (Dietz-Uhler & Hurn, 2013), improve teaching methods, and adopt adaptive learning approaches

(Hu et al., 2014). Similarly, learning analytics create opportunities for instructors and researchers to test new strategies to help students succeed (Niemi & Gitin, 2012). Outside of the student-teacher interaction, students can explore the learning data to self-regulate their own learning strategies, and reflect on their own learning (Xing, Guo, Petakovic, & Goggins, 2015).

References

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

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. http://www.ncolr.org/jiol/issues/pdf/12.1.2.pdf

Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, *69*, 897-904. http://dx.doi.org/10.1016/j.jbusres.2015.07.001

Fiaidhi, J. (2014). The next step for learning analytics. IT Professional; *IEEE Computer Society, 16*(5), 4-8. http://dx.doi.org.libproxy.uoregon.edu/10.1109/MITP.2014.78

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Educational Technology & Society*, *15*(3), 42-57. Retrieved from http://www.jstor.org/stable/jeductechsoci.15.3.42

Haynes Stewart, T.L., Clifton, R.A., Daniels, L.M., Perry, R.P., Chipperfield, J.G., & Ruthig, J.C. (2011). Attributional retraining: Reducing the likelihood of failure. *Social Psychology of Education: An International Journal*, *14*(1), 75-92. doi: 10.1007/s11218-010-9130-2

Hu, Y., Lo, C., & Shih, S. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, *36*, 469-478. http://dx.doi.org/10.1016/j.chb.2014.04.002

Hudson, W.E. (2006). Can an early alert excessive absenteeism warning system be effective in retaining freshman students? *Journal of College Student Retention: Research, Theory, & Practice, 7*(3-4), 217-226 doi:10.2190/8TDY-798N-1ACK-9733

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *Internet & Higher Education, 27*. 54-63. doi:10.1016/j.iheduc.2015.06.001

Lizzio, A., & Wilson, K. (2013). Early intervention to support the academic recovery of the first-year students at risk of non-continuation. *Innovations in Education and Teaching International, 50*(2), 109-120. http://dx.doi.org/10.1080/14703297.2012.760867

Macfadyen, L.P., & Dawson, S. (2010). Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2), pp.588-599. doi:10.1016/j.compedu.2009.09.008

Mijatovic, I., Cudanov, M., Jednak, S., & Kadijevich, D.M. (2013). How the usage of learning management systems influences student achievement. *Teaching in Higher Education, 18*(5), 506-517. http://dx.doi.org/10.1080/13562517.2012.753049

Niemi, D., & Gitin, E. (2012). *Using big data to predict student dropouts: Technology affordances for research*. International Association for Development of the Information Society, Paper presented at the International Association for Development of the Information Society (IADIS) International Conference on Cognition and Exploratory Learning in Digital Age (CELDA) (Madrid, Spain, Oct 19-21, 2012). 4 pp.

Phillips, E. D. (2013). Improving advising using technology and data analytics. *Change: The Magazine of Higher Learning, 45*(1), 48-55. http://dx.doi.org/10.1080/00091383.2013.749151

Smith, V.C., Lange, A., & Huston, D.R. (2012). Predictive modeling to forecast student outcomes and drive effective interventions in online community college courses. *Journal of Asynchronous Learning Networks*, *16*(3), 51-61. **http://sloanconsortium.org/jaln/v16n3/predictive-modeling-forecast-student-outcomes-and-drive-effective-interventions-online-co**

Tempelaar, D.T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior, 47*, 157-167. http://dx.doi.org/10.1016/j.chb.2014.05.038

Thille, C., Schneider, E., Kizilcec, R., Piech, C., Halawa, S., & Greene, D.K. (2014). The future of data-enriched assessment. *Research & Practice in Assessment, 9*. 5-16. Retrieved from http://www.rpajournal.com/the-future-of-data-enriched-assessment/

United States Department of Education (2015). *Fact sheet: Focusing higher education on student success*. Retrieved from: http://www.ed.gov/news/press-releases/fact-sheet-focusing-higher-education-student-success

University of Florida, Center for Public Issues Education (2014). *Evaluating information sources*. Retrieved from: http://www.centerpie.com/wp-content/uploads/2014/08/evaluateinfo.pdf

Vander Schee, B.A. (2011). Early intervention: Using assessment to reduce student attrition. *About Campus, 16*(1), 24-26. http://dx.doi.org/10.1002/abc.20051

Wagner, E., & Ice, P. (2012). Data changes everything: Delivering on the promise of learning analytics in higher education. E

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168-181. http://dx.doi.org/10.1016/j.chb.2014.09.034

Zacharis, N.Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *Internet and Higher Education, 27*, 44-53. http://dx.doi.org/10.1016/j.iheduc.2015.05.002