Learning Analytics in Higher Education: An Annotated Bibliography

Current State of Learning Analytics in Higher Education

Drivers, Trends, Benefits, and Challenges

Bichsel, J. (2012). *Analytics in Higher Education: Benefits, Barriers, Progress, and Recommendations*. EDUCAUSE Center for Applied Research.

The purpose of this study was to develop an analytics maturity model through which institutions can assess their own progress in the use of academic and learning analytics. More than half of the institutions surveyed are actively engaging in activities that meet the definition of data analytics in the areas of enrollment, retention, institutional resource optimization, and financial management. However, most institutions are below level four on the five-point maturity index. Although top leadership shows significant interest in analytics adequate funding is a barrier. Most institutions also scored low on the maturity index for data analytics tools, reporting, and most importantly, according to the author, expertise. However, the study found that institutional leaders who view analytics as an investment in institutional capability are making greater progress with their analytics projects and they are more likely to use data proactively to drive specific interventions or other actions. Based on the findings, the author recommends establishing the value of data-based decision making through planning several small winning projects or a single large one in the areas of enrollment management, cost reduction, or resource optimization. Foremost is the development of analytics capabilities aligned with organizational strategy.

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). Retrieved from www.ncolr.org/jilo

The title suggests coverage of analytics from the faculty perspective. However, this paper is more of a literature review/state of the industry report on learning analytics. Topics covered include retention and performance in higher education, the drivers for increased data-driven decision-making, examples of how specific institutions are using learning analytics, tools and data available, etc. The authors reiterate concerns about data privacy, profiling/bias, and the need for clarity and transparency of goals, policies and practices.

Ferguson, Rebecca (2012). Learning analytics: drivers, developments and challenges. *International Journal of Technology Enhanced Learning, 4*(5/6) pp. 304–317.

The author reviews current trends and suggests future directions for learning analytics research and practice. Key drivers examined include increased focus on learner outcomes, economic concerns related to higher education cost and results, and advances in technology, or more specifically, the massive amounts of data generated as a byproduct of digital learning activity. Ferguson describes the integration of social network analysis as a significant development that distinguishes learning analytics from the related research areas of educational data mining and academic analytics. Referring to online learning as both participatory and social, the author outlines several subsets of social learning analytics research grounded in learning science, including: discourse analytics; learner disposition analytics; and content analytics, including recommender systems that filter and deliver content based on tags and ratings supplied by learners. Forthcoming challenges noted by author in the nascent field of learning analytics include working with more complex datasets containing multiple data types such as mobile, biometric, and learner disposition data the development of a an ethical framework for ownership, stewardship, and usage of learner data.

Siemens, G. & Long, P. (2011, September/October). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, *46*(5), 30–32. Retrieved July 10, 2013 from: http://net.educause.edu/ir/library/pdf/ERM1151.pdf

The education sector's utilization of all the data available for timelier decision-making via real-time evidence and data lag behind that of the business sector. The authors recommend a five-level framework for understanding academic and learning analytics to improve the effectiveness of university decision-making, operations, processes, resource allocation and learning outcomes:

1. Data trails generated by the learner at the course level;
2. Predictive analytics using educational data-mining for learner modeling, clustering and pattern mining;
3. Use of data to inform curriculum development and delivery;
4. Adaptive content delivery and sequencing based on learner behaviors, clusters and models; and
5. Adaptive learning, encompassing not just content, but also engagement, relationships/interactions, and learner supports/scaffolding.

U.S. Department of Education, Office of Educational Technology. (2012). *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*. Washington, DC. Retrieved June 10, 2013, from http://www.ed.gov/edblogs/technology/files/2012/03/edm-la-brief.pdf

This learning analytics brief informs policy makers and administrators how higher education institutions are using analytics to improve student services, grades, retention, and learning outcomes, consistent with the U.S. Department of Education's National Education Technology Plan for 21st century learning. Data mining and analytics, including user modeling, user profiling and clustering, domain modeling, relationship mining and data visualization promise to unveil outcome-oriented actionable insights from specific learning behaviors. Learning providers can use those insights to program real-time automated responses and to inform other student support responses such as adjusting teaching or providing intervention services, as illustrated in the examples provided from commercial and public entities serving the education sector.

Despite the acknowledged challenges of implementing data mining and learning analytics, including the high cost of collection, storage, development of algorithms, interoperable administrative and learning systems (systems/data types), the report recommends combining the data types, with acceptable validity, privacy, and ethical standards applied, for improved predictive power.

<u>Application Concepts, Examples, and Emerging Best Practices</u>

Analytics Platforms/Ecosystems

Forsythe, Jr., R.G., Chacon, F.J., Spicer, D.Z., & Valbuena, A. (2012). Two case studies of learner analytics in the university system of Maryland. *Educause Review Online*. Retrieved from http://www.educause.edu/ero/article/two-case-studies-learner-analytics-university-system-maryland

This article highlights nuanced, but important structural considerations for designing data analytics programs to effectively support university retention efforts and available resources. Two University System of Maryland campuses – Bowie State University (BSU) and University of Maryland Eastern Shore (UMES) are profiled. Both serve large numbers of first-generation college students, often from disadvantaged academic backgrounds and both utilize expanded support networks in their retention efforts.

For BSU, second year retention rates were declining, pinpointing issues in transitioning students from general education requirements to upper level courses in the degree programs. The university adopted a student-centered intervention approach, using Starfish Early Alert/Connect to capture relevant student data. University-developed middleware extracts the student demographics, socioeconomic profiles, and academic performance information from Peoplesoft and class rosters, community group membership, and formative grades from Blackboard Learn. Flags, which are assigned to the profiles of students, can be generated automatically and manually and communicated directly to students and/or the myriad of institutional personnel associated with the students, via the local Starfish client interface, which is customized to the specific end-user groups, and integrated with campus e-mail and calendaring. The Student Success Monitoring System (SSMS); consolidates critical information, saving instructors and support staff the significant effort and time that would be needed otherwise to gather relevant information on a student manually from Peoplesoft and the LMS.

UMES addressed the challenge by adopting a flag-centered approach. In this model, support services tailored to behaviors known to lead to attrition are developed and assigned to specific staff in advance. UMES leveraged an existing investment in Microsoft SQL Server and Microsoft PerformancePoint to extract relevant data from various campus databases and to create dashboards customized to end-user roles. Flags (indicators) are triggered based on those behaviors and the lists are fed into Starfish Early Alert. Staff members responsible for delivering a specific service designed to resolve a given, use that system to monitor their assigned indicators. Using the live dashboard, they are able to retrieve the list of any students needed assistance in that area and delivery them, without waiting for advisor referrals. Though it is too early to determine the actually impact on graduation rates, UMES is reporting a 6% increase in third and fourth year retention rates.

Phillips, E. D. (2013). Improving Advising Using Technology and Data Analytics. *Change: The Magazine of Higher Learning*, *45*(1), 48–55. Retrieved from http://www.changemag.org/Archives/Back_Issues/2013/January-February_2013/improving-advising-full.html

The article describes how Arizona State University (ASU) increased its freshmen retention rate by eight percent over five years, resulting in $1.7 million recurring increased revenues and a 10 percent increase in graduation rates using technology and data analytics to power increased placement precision, identify students at-risk and provide tailored interventions. Using student data, the degree paths, and incoming student information, the system profiles and predicts success and recommends the optimal course sequence for each individual student. Critical courses that are diagnostic of student success in a given major are also locked into the degree pathways early and the system flags students for additional support and advising if they do poorly in those classes.

ASU subsequently launched its Student 360 and retention dashboard. Student 360 consolidates student academic and demographic records along from numerous campus databases into one place. The retention dashboard includes college readiness score (based on students' entrance information), financial aid status, and current academic status, which is provided by faculty, along with recommendations for the students, if applicable, at two points each semester. Advisors can more easily monitor the status of their assigned student caseload (avg. 350 per advisor) and focus attention where it is needed most.

Siemens, G, Gasevic, D., Haythornthwaite, C., Dawson, S. Shum, S.B., Ferguson, R.,…Ryan S.J.d. (2011). *Open Learning Analytics: An Integrated & Modularized Platform*. Retrieved from: http://solaresearch.org/OpenLearningAnalytics.pdf

The authors, who are at the forefront of learning analytics research, propose development of an integrated, yet modular platform or analytics ecosystem with specific standards for analytics plug-ins or extensions and adaptive content. Through predictive modeling, the analytics "ecosystem" would utilize predictive modeling and an early warning system to register changes in key risk behaviors and trigger an automated or human intervention. Essential to the platform are student and instructor-facing dashboards featuring for data visualization and reporting. Discourse analytics, lifelong learner profiles, learner dispositions, social network analytics (cluster reports), and knowledge maps are among the specific metrics recommended. Broad, multi-sourced analytics that are contextual and aligned to institutional teaching and learning outcomes are key to the effectiveness of such a platform, according to the authors.

Data Mining and Predictive Modeling

Barber, R. & Sharkey, M. (2012). Course correction: using analytics to predict course success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 259–262).

This paper describes the data that was mined and combined through two predictive model prototype pilots at University of Phoenix. The data, variables, and modeling methods used were informed by a critical review of the literature and consistent with the approaches used by other universities engaged in predictive analytics. What differentiates this article from those describing similar efforts at other universities is focus and goal articulated by UOP, which is to identify students who are at risk of failing their current course and to produce accurate, timely and actionable information that serves the needs of the academic advisors. The authors emphasize the need for complete data sets, finding that logistical regression dropped incomplete records, resulting in significant data loss. They also acknowledge the significant time and manual resources needed to compile data from separate databases.

Among the key findings: Identification of low and high risk students at the start of the course proved accurate about 90% through the end of the course, but the number of students who fall neither in the high nor low risk categories was significantly high until weeks 2-3. Military status was not predictive at the bachelor or associate level. At the master's degree level, military status was negatively correlated to successful completion. Method of payment or financial aid was not predictive, but status (current or not) of payment was. The ratio of credits attempted to credits earned and cumulative points earned were significant predictors. The second iteration of the model, which included the additional variables proved 85% accurate for all students on the first day of class, increasing to 95% accuracy by the third week. Plans for a third iteration of the pilot include the addition of three additional variables -- major, elapsed time since last course, and participation in an orientation – and examination the predictive accuracy of support vector machines and random forest models.

Ice, P., Díaz, S., Swan, K., Burgess, M., Sharkey, M., Sherrill, J., ... & Okimoto, H. (2012). The PAR Framework proof of concept: Initial findings from a multi-institutional analysis of federated postsecondary data. *Journal of Asynchronous Learning Networks, 16*(3). Retrieved from: http://jaln.sloanconsortium.org/index.php/jaln/article/view/277

Funded by the Gates Foundation with project management provided by WCET, the initial study created a federated dataset of 640,000 student records and 3 million course records from six universities representing community colleges, for profit, and public and private four-year universities. Thirty-three common variables were established and applied to the dataset to develop and test models that could predict retention and completion. The research team examined the relative benefits of different statistical methods in predictive analytics for learning. A logistic regression model was found to predict whether a student would become inactive (a predictor negatively correlated to retention) 72.4%, or remain active 85.1% of the time, thereby providing academic administrators with the information required to apply timely intervention resources where they are most needed. The initial findings also revealed that number of transfer credits was highly associated students remaining in active status (i.e., not dropping out) and that credit hours completed was the most significant predictor of active (vs. inactive) status. CHAID analysis, which provides more visually intuitive output, showed the same, but proved better at identifying interrelationships among the variables, including how the larger trends manifest for subgroups.

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).

The authors describe the initial research leading to the 2010 launch of RioPACE (Progress And Course Engagement), a predictive analytics-based early-alert system for at-risk students. Using the Naïve Bayes classification method, they found that LMS activity, measured by weighted variables such login frequency and course site engagement along with student pace (ahead, on-time, behind schedule) were reliably valid predictors of course success at all points throughout a course. Macfadyen, & Dawson's research (2010, as cited by the authors) corroborates the relationship between LMS activity and course outcomes. The authors also found that non-LMS student data, specifically, concurrent credit load, ratio of credits attempted to credits not-earned and total number of credits accumulated also serve as predictors of course success, though to a lesser extent than the LMS data. Interventions were left to the separate academic disciplines to develop and implement. Most plans involved faculty or staff outreach those students signaled by the system via telephone. Only one-third of the calls resulted in direct contact with the students and there was no statistically significant improvement in the overall success rates. Another intervention yielded an increase in early LMS activity and 40% reduction in course drops among a small sample of students who received automated reminder emails the day before class started. Attempts to scale the initiative have not produced similar results, leaving the authors to conclude that additional intervention research is needed.

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Educational Technology & Society, 15*(3), 133–148.

The authors advance the research on validity of recommendation algorithms by using large public datasets collected from Mendeley, MACE, Organic.Edunet, and other similar entities as part of the first European dataTEL challenge. The authors' goal was to use large datasets containing learner usage data in order to assess the applicability of social recommendation for teaching/learning resources. The key characteristics of the datasets , which are among the first that are both publicly available and representative of learner interactions with online tools and resources, are discussed relative to the various collaborative filtering approaches applied, including Cosine similarity, Pearson correlation, and Tanimoto-Jaccard coefficient. The authors found that the accuracy of the similarity measures varies based on the characteristics and contextual nature of the dataset. For example, Tanimoto was found to be most accurate on sparse datasets. Additionally, the best choice of inferred measures, whether user-based, item-based or slope-one, is also dataset dependent. Finally, when explicit feedback ratings are lacking in the dataset, implicit feedback in the form of captured user interaction with the data, such as downloads and user-assigned tagging proved useful in improving the perceived usefulness of each recommendation (precision) and the number of useful recommendations provided (recall). The authors recommend careful selection and testing of algorithms on comparable large datasets to ensure there is a contextual match to the actual learning situation in which the recommendation application will be used.

Maintaining Academic Engagement and Learning Momentum

Student-Facing Dashboards and Instructor Engagement in Retention Efforts

Arnold, K. E. & Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (pp. 267–270).

This article examines the academic results of the Course Signals predictive data analytics and early intervention project at Purdue University. Developed in 2007, Course Signals can be run at scheduled intervals or on-demand by instructors at the course level (for select courses) to identify and generate learning risk-level reports on the students in the course. Risk status is based on students' academic history/preparation, demographic information, current course performance based on grades, and interaction, relative to peers, with the learning management system, which represents level of effort and engagement. Students receive performance status alerts via a graphic interface resembling a traffic light, and instructors use the intervention report to push specific messages to the students containing information about their learning behavior and performance and where/how the student can obtain help via campus resources or instructor office hours.

The authors believe this approach is consistent with Tinto's research (1993) indicating that student success and intervention program aimed at improving persistence should foster student integration and learning. The student performance date (2007-2009) provided shows a 10.37% increase in As and Bs awarded over previous semesters of the same courses not using Course Signals. The article provides student data showing significantly higher retention rates for students who took at least one course using Course Signals, relative to peers not using Course Signals. The retention rates remained consistently higher for students who take multiple courses using Course Signals, than students who took one or no courses with Course Signals. The impact, according to the authors, is especially pronounced for students with early and multiple occurrences of courses utilizing Course Signals.

Fritz, J. (2013). *Using Analytics at UMBC: Encouraging Student Responsibility and Identifying Effective Course Designs* (Research Bulletin). Louisville, CO: EDUCAUSE Center for Analysis and Research. Retrieved from http://www.educause.edu/ecar

Fritz discusses "Check My Activity" (CMA) a student-facing data reporting tool created and piloted by UMBC in 2008. This initial foray into learning analytics was prompted by a correlation UMBC discovered the previous year that showed 40% less activity/usage of the LMS among those earning grades of D and F than other students. The CMA dashboard report serves as an engagement report of sorts, by allowing students to view their weekly clickstream (hit & login) traffic from a given course in Blackboard relative to a de-identified list of their classmates. The data table includes a link to a grade report for the week showing student progress c compared to the average for the class. Initial findings are that students who use CMA were 1.92 times more likely to earn a grade of C or higher.

Fritz describes CMA's relative feedback approach as a scalable intervention aimed at increasing student performance awareness and motivation, indicates that UMBC will be using the analytics module of Blackboard rather than continuing/maintaining development of CMA. He summarizes survey results from pre-identified at-risk first-year students, indicating that the dashboard data was nearly as effective as human interventions in terms of usage and pass rates, even though students perceived it as less important than the other sources. Fritz also describes another finding uncovered by UMBC during their review of LMS data suggesting a correlation between higher grades the use of Blackboard's adaptive release, which sets pre-conditions students must meet before they are able to access the next content item in the course sequence.

Beyond Alerts: Promoting Success through Personal Informatics and Recommender Systems

Bramucci, R. & Gaston, J. (2012). Sherpa: increasing student success with a recommendation engine. *In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 82–83).

The authors advocate the adaptation of automated recommendation engines (i.e., Amazon, Netflix) in higher education as a way to increase student success at a lower cost. By way of example, they share information about Sherpa, an academic recommendation system that relies on predictive analytics and inference rules coded by subject matter experts (SMEs) to assist the 43,0000 students in the South Orange County Community College District (SOCCCD) in making better academic decisions. The recommendation engine delivers personalized messages to students in a variety of formats, including email, SMS, voice, and Facebook announcements, based on specific time and event triggers. The system integrates and expands on SOCCCD's "MySite," the enterprise academic web portal, and "My Academic Plan (MAP)," the online course planning tool used by students since 2007. Students receive information based on a match between the student's personal attributes and the target population attributes provided by the information author and/or whether the student has opted-in to different feeds available via the college MySite portal.

Information may include recommendations for acceptable alternative courses when courses are full, campus events, links online services such as course registration, textbook ordering, or matriculation, or print, audio or video nudges crafted by SMEs for specific target audiences using Boolean operators. Nudges are triggered by dates and actions and actions are triggered by data changes. Students can also rate and comment on the nudges, which provides feedback to the system on the perceived importance or value of the nudge and the clarity of the message. Future plans for the system include adding datamining capability to enable course recommendations based on successful completion by students with similar profiles and adding optional location-based recommendations using mobile phone GPS services.

Frankfort, J., Salim, K., Carmean, C., & Haynie, T. (2012). Analytics, nudges, and learner persistence. *Educause Review Online.* Retrieved from www.educause.edu/ero/article/analytics-nudges-and-learner-persistence

The authors share preliminary non-quantified results and anecdotal evidence to show that students who used the Persistence Plus performed better than those who did not. The two different introductory math courses in the pilot were University of Washington Tacoma's first online course offerings. Persistence Plus currently uses a combination of human/artificial intelligence and manual/automated processes to provide students with behavioral modification/reinforcement and intervention support via static and interactive messages delivered to students' mobile devices. Persistence Plus relies on information students are prompted to enter at regular intervals about their moods, beliefs, mindsets, goals, upcoming assignments/assessments, and grades. The application then sends text messages, in the form of nudges, questions, and prompts, tailored and timed to that input. Persistence Plus messages are intended to encourage students to engage in behaviors, which based on learning and behavioral research, have been strongly correlated with goal attainment, persistence and retention. Support messages may also include encouragement, information about free campus tutoring, advisor office hours links to Web resources, advice from recent graduates, or other self-help resources. Data on student interaction /engagement with the application serves as an early warning indicator to instructors or advisors that direct human intervention and support may be needed. One quantifiable result shared by the authors, demonstrated student willingness to interact with the application: More than 85% of pilot participants responded to at one question. Some students have also entered appreciation messages in response to the support nudges.

Duval, E. (2011). Attention please! Learning analytics for visualization and recommendation. In Proceedings of LAK11: 1st International Conference on Learning Analytics and Knowledge (pp. 9–17).

Duval brings a somewhat different perspective to the mix, framing the problem as one of competing demands on attention and referring to click traffic as indicative of engagement as attention metadata. The author draws his research inspiration from what others have dubbed "the quantified self-movement," or personal informatics, the goal of which is to actively engage individuals in the use of tracking tools (i.e., www.rescuetime.com and http://runkeeper.com/) to increase awareness and understanding, through visualization of quantitative data, the interplay of time and behavior on task accomplishment or goal attainment. Tracking data aggregated from many users can be used to display relative performance and drive automated next action recommendations based on similar clusters of end-users and even and provide motivational nudges. By way of educational example, Duval references several preliminary pilot studies on tracking widgets that can be embedded in an LMS or other type of personal or online learning environment to support teaching and learning through activity visualization. He further advocates the development of systems that use learner attention metadata to recommend specific resources, activities or people (experts or others in the learner's social network), based on specific learner objectives. While acknowledging the need for more research on the behavioral impact special purpose networks, the author points to a fitness related study by (Ma, Chen, & Xiao, 2010, November) suggesting that reporting/sharing of tracking information with a social network of like-minded peers influenced goal attainment. He also references research by (Swan, 2009) showing that the use of personal informatics and social networks may result in more patient-driven approaches to health care. Finally, Duval suggests that the ability to deploy and scale such systems will depend on open global interoperability standards and protocols.

Social Learning Analytics

Ferguson, R. & Shum, S. B. (2012). Social learning analytics: five approaches. *In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 23–33).

This paper serves as a well-researched primer on the development and current applications of Social Learning Analytics in online learning. Among the catalysts identified by the authors as driving the increase in social learning are:

- The explosion of social media, driven in part by readily available internet access, including via mobile device, in affluent countries;
- The increasing value placed on collaboration as a 21st century workplace skill, particularly as it relates to effective innovation;
- The growing amount of content available via the Web and the corresponding expectation by learners that said content will be free and of reasonable quality;
- The need for managing one's attention and time while filtering and navigating the vast amount of information available as a result of the Internet

The authors identify and describe five dimensions along with social learning can be analyzed: Social Networks, Discourse, Learner Dispositions (not to be confused with learning styles), Content, and Context. While the first two are inherently social, the latter three are becoming increasingly socialized in the context of online learning. They also describe the application of these analytics on SocialLearn, an investigative social media learning platform used by the Open University and the ways learners interact with them, both as data sources and as recipients of the information, often in the form of visualizations, provided by the analytics.

Sie, R. L., Ullmann, T. D., Rajagopal, K., Cela, K., Rijpkema, M. B. & Sloep, P. B. (2012). Social network analysis for technology-enhanced learning: review and future directions. *International Journal of Technology Enhanced Learning, 4*(3/4), 172–190.

A review of the literature to date, with the goal of framing the potential value of social network analysis (SNA) in online learning. Sieman's (2005) social learning theory of connectivism is discussed as are the two primary frameworks for data connection and analysis – sociocentric and egocentric (Garton, Haythornthwaite, & Wellman, 1997 and Chung & Davis, 2005). The former addresses the structure of member interactions at the network level, whereas the latter identifies the linked contacts of individuals. Key network metrics include the ratio of existing links to total number of links possible (density), which at the individual level can reveal how well the members know each other; the level of centralization/decentralization as an indicator of dependencies within the network; and clustering within the network, which reveals the number of communities of practice or interest that exist within a social network. The authors also outline the primary techniques used for network analysis, including network visualization, network analysis (of the visualization), simulation and network intervention and summarize the research findings on each.

According to the authors, the primary value of visualization and analysis is in informing social interventions, which may include decreasing centralization (over-dependence on a few individuals or resources), increasing the number of learners or connections among each learner; strengthening ties between types of learners, and supporting in) individual learners in the network, especially those the network visualization reveals as disconnected. The authors suggest that the data trails afforded by LMS environments are well suited to network analysis and the development of network based recommender systems. However, more research is needed using two-mode visualizations to better understand the content of the relationships within learning networks (i.e., what learners are talking about) and the relationships in between learners and learning resources, objects, and tools (i.e., how learners are learning).

References

Chung, K., & Davis, J. (2005). Exploring sociocentric and egocentric approaches for social network analysis. *Knowledge Management in Asia Pacific (KMAP 2005)*, (pp. 1-8). Wellington, New Zealand.

Garton, L., Haythornthwaite, C., & Wellman, B. (1997). Studying online social networks. *Journal of Computer-Mediated Communcation, 3*(1), 1-29.

Ma, X., Chen, G., & Xiao, J. (2010, November). Analysis of An Online Health Social Network. *In Proceedings of the 1st ACM International Health Informatics Symposium* (pp. 297–306). ACM.

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), 588–599.

Siemans, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning, 2*(1), 3–10.

Swan, M. (2009, February). Emgerging patient-driven health care models: An examination of health social networks, consumer personalized medicine and quantified self-tracking. *International Journal of Environmental Research and Public Health, 6*(2), 492-525.

Tinto, V. (1993). *Leaving college: Rethinking the causes and cures of student attrition* (2nd ed.). Chicago: University of Chicago Press.