Chapter 2
Thoughts on Recent Trends and Future Research Perspectives in Big Data and Analytics in Higher Education

Jay Liebowitz

Abstract In many sectors, including education, the growth of data has been increasing dramatically over the years. In order to make sense of this data and improve decision-making, analytics and intuition-based decision-making should be key components in this "Big Data" era. Educational data mining and learning analytics are becoming the lingua franca for those institutions who seek to improve their strategic and operational decision-making abilities. This chapter highlights some thoughts in these areas.

Keywords Learning analytics • Big Data • Educational data mining • Intuition • Decision-making

Big Data Ahead

Many sectors are facing the onslaught of massive amounts, speeds, and varieties of data for organizational consumption. Jagadish et al. (2014) states that:

> We have entered an era of Big Data. Many sectors of our economy are now moving to a data-driven decision making model where the core business relies on analysis of large and diverse volumes of data that are continually being produced. (p. 86)

Data growth in some sectors like healthcare, education, and others are growing as much as 35 % a year (Liebowitz, 2013, 2014a, 2014b). The SAS Institute predicts that there will be a 240 % growth by 2017 for employees needed to handle Big Data tasks. Gartner predicts that by 2017, cloud-empowered chief marketing officers

J. Liebowitz (✉)
Harrisburg University of Science and Technology, 326 Market Street,
Harrisburg, PA 17101, USA
e-mail: jliebowitz@harrisburgu.edu

B. Kei Daniel (ed.), *Big Data and Learning Analytics in Higher Education*,
DOI 10.1007/978-3-319-06520-5_2

(CMOs) will spend more on IT than chief information officers (CIOs) do (Gorenberg, 2014).

We are already seeing a variety of analytics applications targeting the CMO such as optimizing cross-channel influences between online and offline marketing (e.g., OptiMine) and increasing mobile app users' loyalty and conversion rate tied to corporate customer relationship management (e.g., FollowAnalytics) (Gorenberg, 2014). According to the 2014 Big Data in Retail Study, commissioned by 1010 data, 96 % of the respondents, all of them executives in retailing, think Big Data is important in keeping retailers competitive.

No matter where you turn, Big Data will have an impact. The education sector is no different. Lane (2014) points this out in his edited book about how to build a smarter university through Big Data, Analytics, and Innovation. Whether looking at student success measures, student retention, learning outcomes, or other educational metrics, the use of analytics and predictive modeling can be an asset to navigate through these Big Data waters. The Aberdeen Research Group, in their 2014 business intelligence (BI) in Retail Industry survey of over 200 retail companies, showed that to move from Industry Average to Best-in-Class, the following actions are recommended: (1) implement data integration and cleansing tools, (2) aggregate and clean your data, (3) institute BI application development procedures, and (4) create a continued education training program for BI users (in fact, many organizations have a BI competency center (BICC) for this purpose).

In the same research study, in order to improve as a best-in-class company, the following should be done: (1) improve exception reporting, (2) adopt scorecards to measure and track performance, and (3) update all data, dashboards, and reports in real-time across all channels. In many ways, universities and colleges can apply these suggestions in their own "business." In the following paragraphs, we will identify some important Big Data and Analytics techniques, trends, and research issues as related to the education sector. Hopefully, this will provide a platform for further investigation into these areas.

BI/Analytics Conceptual Framework

Before delving into some targeted areas for future development, it may be helpful to develop a BI/analytics conceptual framework to allow students and faculty to further test and validate. To date, there have been very few, if any, conceptual frameworks in the BI/analytics area—no matter what sector is explored. Liebowitz (2014c), based on his research and best practices in the BI/Analytics area, proposes the following conceptual framework as shown in Fig. 2.1 for further testing and enhancement.

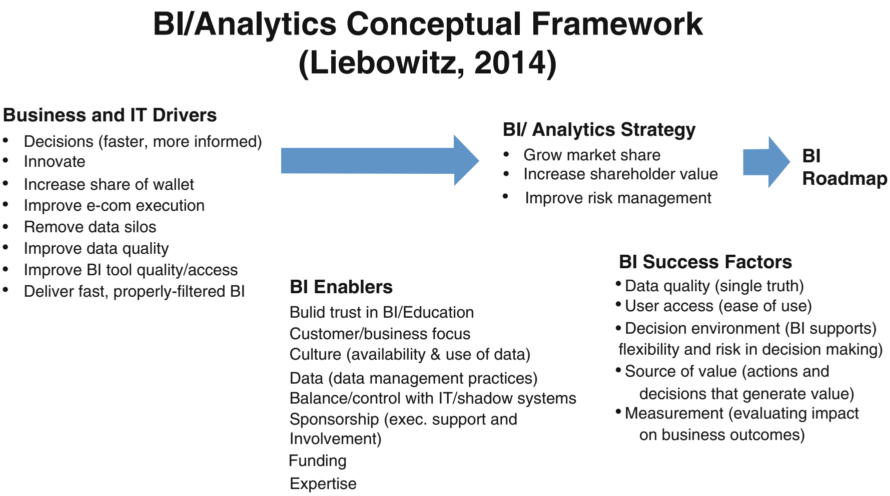

Fig. 2.1 BI/analytics conceptual framework (Liebowitz, 2014c)

The framework shows that there are business and IT drivers that influence the BI/analytics strategy of the organization. There are also BI enablers that impact how successful is the BI/Analytics strategy. As part of the framework, there are BI Success Factors that can be used to derive value from the BI/Analytics strategy. Last, a BI road map (typically 3 years) is built based upon the strategy. This conceptual framework is fairly generic across industries, although there may be some different factors in each area.

Adaptive Learning/Courseware and Educational Data Mining

Adaptive learning/courseware and educational data mining are interesting areas that continue to evolve over the recent years (Chen, Chiang, & Storey, 2012; Siemens & Baker, 2012). Adaptive learning is often equated with personalized e-learning, adaptive courseware, and intelligent tutoring systems. Grubisic (2013) presents a fairly recent review of the literature in these areas. In reviewing the literature, adaptive learning as pertains to e-learning often involves the WHERE-WHY-WHAT-HOW for adaptation—that is, the adaptive system (WHERE), adapting goals (WHY), focus of adaptation (WHAT), and adaptation methods and techniques (HOW). Grubisic (2013) found 5924 papers that relate to either adaptive e-learning systems, intelligent tutoring systems, courseware generation, courseware sequencing, automatic courseware, dynamic courseware, adaptive courseware, or automatic generation of courseware. Twenty-one percent (21 %) of the papers are related to adaptive e-learning systems. For adaptive e-learning systems, two main approaches are used to achieve adaptability (macro- and micro-adaptation). The macro-adaptation occurs before the learning and teaching process, whereby data about a student's cognitive abilities are collected, and then this information drives the type of learning environment and instruction that will best suit those decisions (Grubisic, 2013).

Micro-adaptation occurs during the learning and teaching process, whereby the learning approaches vary according to the knowledge of the student user during the e-learning session. This latter approach may also be called "personalized learning." According to Garrido and Onaindia (2013), some of the challenges of e-learning include selecting the proper learning objects, defining their relationships, and adapting their sequencing to the student's specific needs, objectives, and background. Garrido and Onaindia (2013) developed an approach for assembling learning objects for personalized learning through artificial intelligence (AI) planning. They apply metadata labeling and an AI planning/scheduling technical mapping methodology for course generation.

Educational data mining is related to adaptive learning and is focused on uncovering hidden patterns and relationships to drive student learning outcomes (Romero, Ventura, Pechenizkiy, & Baker, 2011). For example, the University of Maryland University College (UMUC), Kresge Foundation grant applies data mining to identify the critical "at-risk" transfer students in the first or second semester of their UMUC programs in order to provide the necessary support services to help them succeed toward graduation, as explained shortly. Pena-Ayala (2013) performed a recent review of the educational data mining (EDM) field by examining 240 EDM papers from 2010 through the first quarter of 2013.

Usually, descriptive or predictive approaches are applied to EDM. In terms of future trends, EDM modules may become more integrated within the typical architecture of an educational system. Also, the educational environment must continue to advance the notion of data-based decision-making, which highlights the importance of Big Data and Analytics in an educational environment. Last, from a technology viewpoint, EDM will be enhanced by advances in social networks, web and text mining, virtual 3-D environments, spatial mining, semantic mining, collaborative learning, Big Data architectures, and other technology areas (Pena-Ayala, 2013).

Example: Data Mining and Data Integration: A Community College and University Partnership to Improve Transfer Student Success Funded by the Kresge Foundation (Nadasen, 2013)

An interesting 3-year, $1.2 million Kresge educational data mining grant was pursued at the University of Maryland University College (UMUC). The UMUC, Prince George's Community College, and Montgomery College collaborated to build an integrated database to make data-driven decisions on how to improve student success. The students are working adults who enrolled at a community college, then transferred to UMUC. Data mining techniques and statistical analyses were used to analyze the integrated data to identify relationships among variables.

The research objectives were:

- Prepare an integrated database with collaborative community college partners.
- Identify success and failure factors using data mining techniques.
- Build predictive models using statistical analysis and results from data mining.
- Build student profiles and implement models for decision-making.
- Track the impact on retention and graduation.

UMUC is an online institution that enrolls over 90,000 students each year worldwide. The Prince George's Community College (PGCC) is located within 2 miles of UMUC's Academic Center and transfers over 37,000 students. The Montgomery College (MC) is located within 10 miles of UMUC's largest regional center and enrolls over 35,000 students. Through the collaborative data-sharing process, the integrated database contains information on over 11,000 students who transferred from PGCC to UMUC and over 10,000 students transferred from MC. The Data Exchange was set up as follows:

- A memorandum of understanding (MOU) governed the data-sharing agreement and the protection of individual student data. The data were restricted to four researchers who were funded by the Kresge Foundation and UMUC.
- Historic course information on students identified as UMUC transfer students were provided by the community colleges using a secured transfer process.
- Data were stored in a UMUC database located on an Oracle Exadata machine.
- The database contains millions of records on student demographics, courses, and online classroom activities.

Figures 2.2, 2.3, and 2.4 show the tools used in this educational data mining application, as well as the interactions and datasets used.

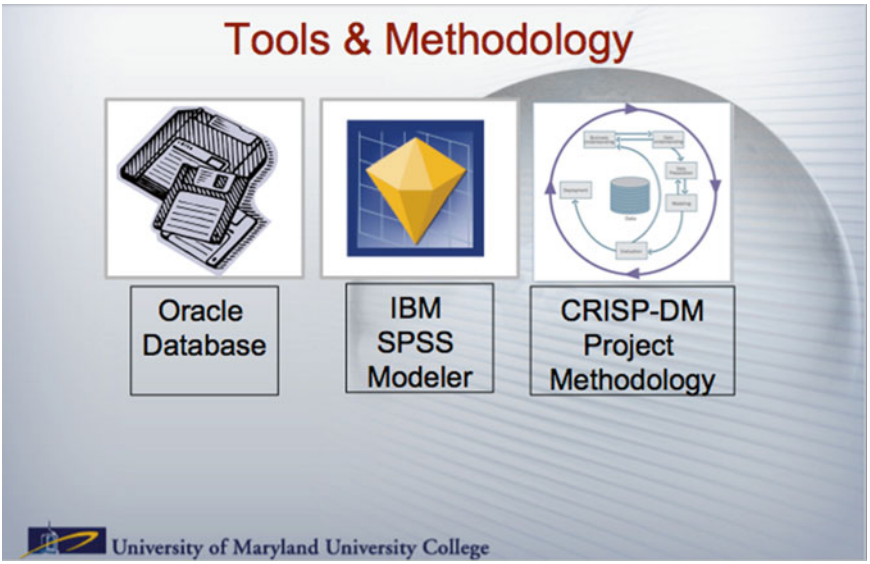

Fig. 2.2 Tools and methodology

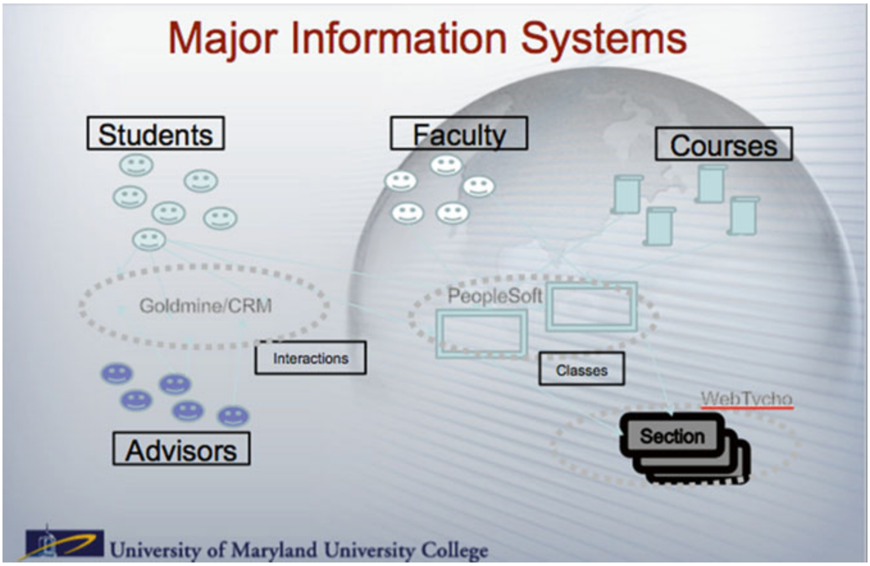

Fig. 2.3 Major information systems used (note: WebTycho was our learning management system (LMS))

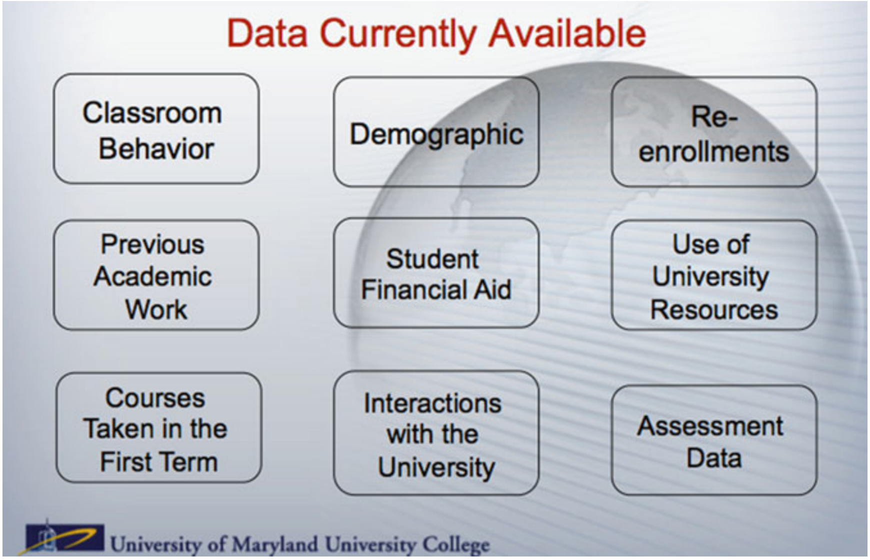

Fig. 2.4 Data currently available in the data mining application

Selected results from this data mining application showed that course-taking behavior prior to transfer influences success at the subsequent institution. Also, online classroom activity prior to the first day of class can predict course success.

Faculty engagement is critical for course/student success. A Civitas pilot, with our data, predicted with 85 % confidence how successful a transfer student would be in their UMUC program within 8 days of starting their program (Nadasen, 2013).

Data Visualization and Visual Analytics

The old adage still holds true—a picture is worth a thousand words. Especially, for C-level executives, data visualization is paramount for understandability and a quick grasp of the analytics/Big Data results. No matter whether it's a CEO of a company or a university president, the analytics results must be displayed in a manner that is easy to understand. Executive dashboards (such as through the popular Tableau Software) are one way to apply data visualization on the analytics results.

One of the key areas of growth in the educational Big Data/analytics field is the use of visual analytics. SAS, probably the leading tool vendor for analytics on the market worldwide, has SAS Visual Analytics as a fairly new feature within its software toolset. Figure 2.5 shows an example of a student analysis using SAS Visual Analytics (of course, there is an interactive component which is hard to see from this static figure).

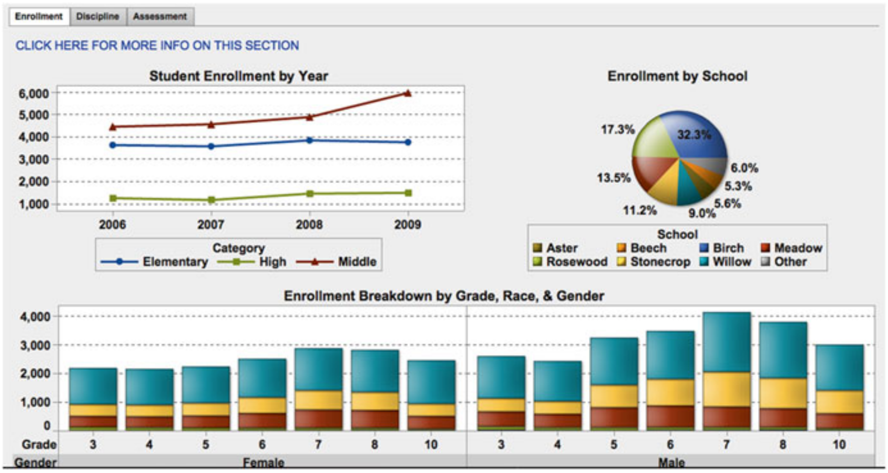

Fig. 2.5 K-12 student analysis using SAS Visual Analytics [http://www.sas.com/software/visual-analytics/demos/k12-student-analysis.html]

Figure 2.6 shows an example using Tableau Software (www.tableau.com) in terms of a district-level evaluation dashboard comparing student scores with meal plans over time.

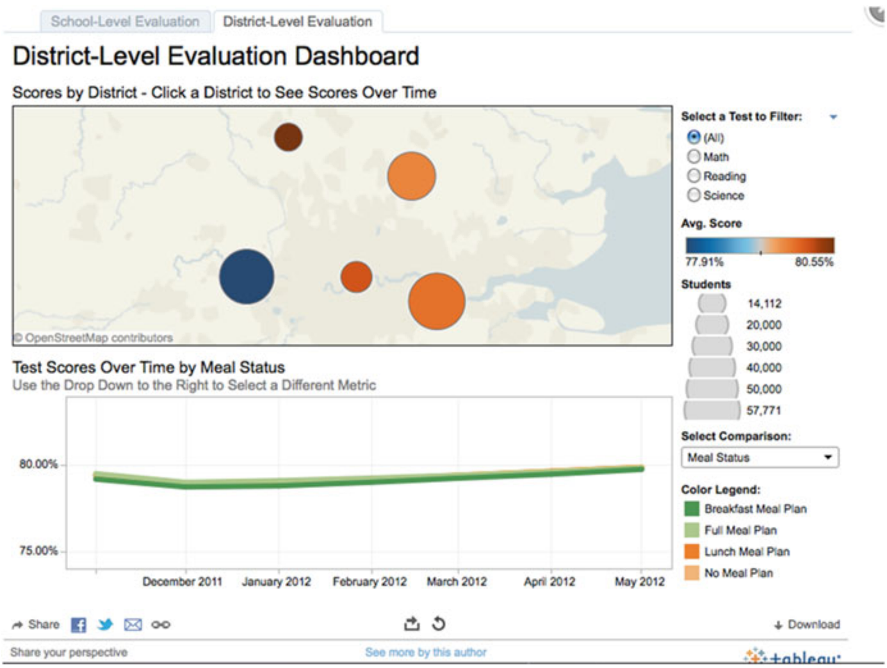

Fig. 2.6 District-level evaluation dashboard using Tableau Software [http://www.tableausoftware.com/solutions/education-analytics]

Over the years, there have been many research agendas suggested for the visual analytics field. For example, Thomas and Cook (2006) recommended the following in terms of moving visual analytics from research into practice:

- Facilitate understanding of massive and continually growing collections of data of multiple types.
- Provide frameworks for analyzing spatial and temporal data.
- Support the understanding of uncertain, incomplete, and often misleading information.
- Provide user- and task-adaptable guided representations that enable full situation awareness while supporting development of detailed actions.
- Support multiple levels of data and information abstraction, including integration of different types of information into a single representation.

- To accelerate the transition of research into analytical practice, the R&D community must:
  - Develop an infrastructure to facilitate evaluation of new visual analytics technologies.
  - Create and use a common security and privacy infrastructure, incorporating privacy-supporting technologies such as data minimization and data anonymization.
  - Use a component-based software development approach for visual analytics software to facilitate evaluation of research results in integrated prototypes and deployment of promising components in diverse operational environments.
  - Identify and publicize the best practices for inserting visual analytics technologies into operational environments.

Many organizations, such as the Educause and the Knowledge Media Institute (UK), are looking at how visual analytics can be improved in the education environment. For example, Educause's Learning Analytics Initiative has been looking partly at how visual analytics play a key role in education [http://simon.buckinghamshum.net/2012/04/educause-learning-analytics-talk/]. The Catalyst project at the Knowledge Media Institute augments existing social media platforms with web-based annotation tools, recommenders to help users prioritize attention, online creativity triggers, interactive visualizations, and social network and deliberation analytics [http://kmi.open.ac.uk/projects/name/catalyst]. Certainly, in the years ahead, data visualization and visual analytics will continue to evolve and further allow the decision-maker to analyze data in immersive environments and interactive gaming scenarios.

Knowledge Management in Education

Another important area that we will see increased attention deals with knowledge management issues in education, especially as Big Data and analytics continue to evolve. Knowledge management (KM) refers to how best to leverage knowledge internally and externally in an organization (Liebowitz, 2012a, 2012b). More specifically, it deals with knowledge retention and transfer issues, and organizations are applying knowledge management to increase innovation, build the institutional memory of the organization, allow for adaptability and agility, and improve organizational internal and external effectiveness.

In the education field, there are organizations like the Institute for the Study of Knowledge Management in Education (http://www.iskme.org/) whose mission is to improve the practice of continuous learning, collaboration, and change in the education sector. At the Vanderbilt University Medical Center, their Knowledge Management Informatics Center provides high-level data and knowledge organization skills to optimize the enterprise clinical, research, and educational initiatives [http://www.mc.vanderbilt.edu/km/]. The Notre Dame of Maryland University has a new Master of Science in Analytics in Knowledge Management to look at the synergies between knowledge management and analytics [http://www.ndm.edu/academics/school-of-arts-and-sciences/programs/ms-in-knowledge-management/]. We are also seeing a variety of applications of applying knowledge management with e-learning (Liebowitz & Frank, 2010). There is even a relatively new journal from the Faculty of Education in the University of Hong Kong titled *Knowledge Management & E-Learning* [http://www.kmel-journal.org/ojs/].

The 2014 Horizon Report reports short-, mid-, and long-term trends in higher education. One of the six key trends cited is the "Rise of Data-Driven Learning and Assessment," as a midrange trend in driving changes in higher education within 3–5 years [http://cdn.nmc.org/media/2014-nmc-horizon-report-he-EN-SC.pdf]. Here again, the synergy between Big Data, analytics, and knowledge management becomes apparent. For example, the confluence of these areas may produce a strategy to identify "at-risk" students through assessment of critical knowledge areas and competencies via looking at hidden patterns and relationships in large masses of student data and other databases. Coupled with this data-driven approach should be the application of intuition-based decision-making (Liebowitz, 2014b), formed through experiential learning.

Summary

The years ahead look bright for the application of Big Data and analytics in higher education. Certainly, adaptive/personalized learning, educational data mining, data visualization, visual analytics, knowledge management, and blended/e-learning will continue to play growing roles to better inform higher education officials and teachers. And of course, analytics plus intuition should equal success as decision-makers apply "rational intuition" in their education challenges and opportunities.

References

Chen, H. Chiang, R. H. L., & Storey, V. C. (Eds.). (2012, December). Special issue on "business intelligence and analytics: from big data to big impact". *MIS Quarterly*, *36*(4).

Garrido, A., & Onaindia, E. (2013). Assembling learning objects for personalized learning: An AI planning perspective. *IEEE Intelligent Systems., 28*, 64–73.

Gorenberg, M. (2014). Investing in analytics: Optimizing the data economy. IEEE Computer.

Grubisic, A. (2013). Adaptive courseware: A literature review. *Journal of Universal Computer Science, 21*(9), 1168–1209.

Jagadish, H., Gehrke, J., Labrinidis, A., Papakonstantinou, Y., Patel, J., Ramakrishnan, R., et al. (2014). Big data and its technical challenges. *Communications of the ACM, 57*(7), 86–94.

Lane, J. (Ed.). (2014). *Building a smarter university: Big data, innovation, and analytics*. Albany, NY: SUNY Press.

Liebowitz, J. (Ed.). (2012a). *Knowledge management handbook: Collaboration and social networking* (2nd ed.). Boca Raton, FL: CRC Press.

Liebowitz, J. (Ed.). (2012b). *Beyond knowledge management: What every leader should know*. New York: Taylor & Francis.

Liebowitz, J. (Ed.). (2013). *Big data and business analytics*. New York: Taylor & Francis.

Liebowitz, J. (Ed.). (2014a). *Business analytics: An introduction*. New York: Taylor & Francis.

Liebowitz, J. (Ed.). (2014b). *Bursting the big data bubble: The case for intuition-based decision making*. New York: Taylor & Francis.

Liebowitz, J. (2014c). *"Editorial: A conceptual framework for business intelligence/analytics", submitted to INFORMS Analytics*.

Liebowitz, J., & Frank, M. (Eds.). (2010). *Knowledge management and E-learning*. New York: Taylor & Francis.

Nadasen, D. (2013). *"Data mining and data integration: A community college and university partnership to improve transfer student success" summary slides*. Adelphi, MD: University of Maryland University College, Office of Institutional Research.

Pena-Ayala, A. (2013). Educational data mining: A review of recent works and a data mining-based analysis of the state-of-the-art, Expert Systems With Applications: An Int. Journal, Elsevier.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. (Eds.). (2011). *Handbook on educational data mining*. Boca Raton, FL: CRC Press.

Siemens, G., & Baker, R. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd Int. Conference on Learning Analytics and Knowledge, Association for Computing Machinery (ACM)*.

Thomas, J., & Cook, K. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications, 26*(1), 10–13.

http://www.springer.com/978-3-319-06519-9

Big Data and Learning Analytics in Higher Education
Current Theory and Practice
Kei Daniel, B. (Ed.)
2017, XX, 272 p. 56 illus., 48 illus. in color., Hardcover
ISBN: 978-3-319-06519-9