Tech Know Learn (2017) 22:237–241
DOI 10.1007/s10758-017-9331-2

EDITORIAL

Big Data in Higher Education: Research Methods and Analytics Supporting the Learning Journey

David Gibson[1]

Published online: 24 July 2017

1 Introduction

One of the promises of mining big data for insights in higher education is to enable a new level of evidence-based research into learning and teaching. The broader term *data science*, which can be applied to many types and kinds of data big and small, captures a theoretical and methodological sea change occurring in educational and social science research methods that is situated apart from or perhaps between traditional qualitative and quantitative methods (Gibson and Ifenthaler 2017; Gibson and Webb 2015). Today, due to the fine-grained data captured during digital learning, it is possible to gain highly detailed insight into student performance and learning trajectories as required for personalizing and adapting curriculum as well as assessment (Baker and Yacef 2009).

In this new era of data-driven learning and teaching, researchers need to be equipped for the change with an advanced set of competencies that encompass areas needed for computationally intensive research (Buckingham Shum et al. 2013). For example, new data-management techniques are needed for big data, and new knowledge is needed for working with interdisciplinary teams with members who understand programming languages as well as the cognitive, behavioral, social and emotional perspectives on learning. A new horizon of professional knowledge is needed, including new heuristics, which incline a researcher or teaching-researcher toward computational modeling when tackling complex research problems (Gibson 2012).

This special issue on analytics in higher education learning and teaching focuses on some of the enabling computational approaches and challenges in research concerning the journey of a learner from pre-university experiences, recruitment, personalized learning, adaptive curriculum and assessment resources and effective teaching, to post-university

✉ David Gibson
david.c.gibson@curtin.edu.au

1 Curtin University, Bentley, WA, Australia

life-long learning. The collection includes three original research articles, three work-in-progress reports, two articles on emerging technology and one integrative review.

2 Empirical Investigations

Empirical investigations report quantitative or qualitative research demonstrating advances in digital learning, gamification, automated assessment or learning analytics. In this section the three articles address understanding the college-going aspirations of students and their journey into higher education, the design of adaptive learning experiences, and building predictive early warning systems.

Research on factors leading to college-going choices of middle students has not yet utilized the extensive fine-grained data now becoming available on learning and engagement. This situation led the team of Maria Ofelia Z. San Pedro, Ryan S. Baker, and Neil T. Heffernan to collaborate on *An Integrated Look at Middle School Engagement and Learning in Digital Environments as Precursors to College Attendance*. The team used data mining methods on interaction-based assessments of student behavior, academic emotions and knowledge gathered from a middle school online learning environment, and evaluated relationships among the factors for impact on outcomes in high school and college. The measures were used to develop a prediction model of college attendance, to examine relationships concerning intermediate outcomes on the journey to college and to develop a path model for the educational experiences students have during middle school, high school, and college attendance. The research provides a picture of the cognitive and non-cognitive mechanisms that students experience throughout varied phases in their years in school, and how those mechanisms may be related to one another. Such understanding may provide educators with new information about students' trajectories within the college pipeline.

Once students are enrolled into higher education, the issue of delivering a personalized curriculum becomes of prime interest. In *Using Data to Understand How to Better Design Adaptive Learning*, Min Liu, Jina Kang, Wenting Zou, Hyeyeon Lee, Zilong Pan and Stephanie Corliss investigate how the behavior patterns of learners with different characteristics interact with an adaptive learning environment. They collected data from the needs and interests of incoming 1st-year students in a pharmacy professional degree program who were engaged in an adaptive learning intervention that provided remedial instruction. The study found that affective factors such as motivation as well as the alignment among system components had an impact on how learners accessed and performed. Data visualizations revealed relationships that might have been otherwise missed. Their article is part of the bigger picture of how exploratory data mining can help inform the design of adaptive learning environments.

Another issue of great concern is how to make predictions that trigger early interventions that might help prevent attrition. The research project reported in *Predicting Student Success: A Naïve Bayesian Application to Community College Data by* Fermin Ornelas and Carlos Ordonez describes how this team developed and implemented a continuous Naïve Bayesian Classifier for courses at a community college. The method improved the teams' previous prediction accuracy from 70 to 90% for both at-risk and successful students while easing some of the challenges of interpretation of results and implementation of interventions. Predictive results were obtained across eleven courses and cumulative gain charts show the potential for improvements that are made possible by focusing on high-level risk students. The findings may be relevant for implementation of early alert systems in other higher education contexts.

3 Work-in-Progress Studies

Work-in-progress studies provide early insights into leading projects or that document progressions of excellent research within the field of digital learning, gamification, automated assessment or learning analytics. There are three studies in this section that focus on student perceptions of fine-grained analytics on dashboards, an analysis challenge concerning the grain size of analytic observations and the impacts on students of an augmented reality serious game approach to onboarding into the university.

With the increased analytics capability in higher education, more institutions are developing or implementing student dashboards. But despite their emergence, students have had limited involvement in the development process. The ongoing research project reported in *Give me a customizable dashboard: Personalized learning analytics dashboards in higher education* by Lynne D. Roberts, Joel A. Howell, and Kristen Seaman reports on student perceptions of and preferences concerning dashboards. Four focus group transcripts representing 41 students identified five key themes including: 'provide everyone with the same learning opportunities', 'to compare or not to compare', 'dashboard privacy', 'automate alerts' and 'make it meaningful—give me a customizable dashboard'. A content analysis of students' drawings of desired dashboards demonstrates that students are interested in learning opportunities, comparisons to peers and personally meaningful data. A survey of students reported here highlights the tension between students' personal autonomy and the collective uniform activity required to ensure equity. The research suggests potential for providing students with a level of control over their learning analytics as a means to increase self-regulated learning and academic achievement. This evolving research is aimed at better understanding students emotional and behavioral responses to feedback and alerts on dashboards.

'Supplemental Instruction' is a voluntary, non-remedial, peer-facilitated, course-specific intervention that has been widely demonstrated to increase student success, yet concerns persist regarding the biasing effects of disproportionate participation by already higher-performing students. With a focus on maintaining access for all students, the research team of Maureen A. Guarcello, Richard A. Levine, Joshua Beemer, James P. Frazee, Mark A. Laumakis, and Stephen A. Schellenberg examined data from a large, public university in the Western United States. In the article *Balancing Student Success: Assessing Supplemental Instruction through Coarsened Exact Matching* the team used data including student demographics, performance, and participation in supplemental programs to evaluate the efficacy of supplemental instruction. The analysis was conducted in the first year of implementation within a traditionally high-challenge introductory psychology course. Findings indicate a statistically significant relationship between student participation in supplemental instruction and increased odds of successful course completion. Furthermore, the application of Coarsened Exact Matching reduced concerns that increased course performance was attributed to an over-representation of higher performing students who elected to attend the voluntary sessions.

Students in Hong Kong are introduced to academic integrity and ethics issues through mobile Augmented Reality learning trails—Trails of Integrity and Ethics—which are accessed on smart devices. In *Bringing abstract academic integrity and ethical concepts* *into real-life situations*, Theresa Kwong, Eva Wong and Kevin Yue report on the exploratory analytics being conducted on the initial stages of their large-scale, government-funded project which inducts university students. The augmented reality trails immerse students in collaborative problem solving tasks centered on ethical dilemmas, addressed in real locations where various dilemmas might arise, and gives contextually appropriate digital advice and information on demand and as-needed. Students play out the consequences of their decisions, which help reinforce the links between the theoretical concept of academic integrity and ethics and the practical application in everyday contexts. To evaluate the effectiveness of the experiences, the analysis triangulates user experience surveys, qualitative feedback, clickstream data, and text mining of pre- and post- discussions. Preliminary analysis of thousands of student responses suggests that augmented reality learning trails can be adopted and applied to a wider scope of the academic curriculum and co-curriculum.

4 Emerging Technology Reports

The emerging technology reports section presents two views on developments in educational technology that address new potentials for digital learning environments. In *nStudy: A System for Researching Information Problem Solving*, Philip H. Winne, John C. Nesbit and Fred Popowich discuss a new technology for tracing how students work on solving information problems. The platform and toolset gathers fine-grained data about what students do as they work on information problems, which information they work with, and how they adapt tactics and strategies in response to feedback. The technology is implemented as an extension to the Chrome web browser supported by a server-side database that warehouses logged trace data. Trace data record the information learners operate on and operations they apply to that information. Peripheral systems on the server extract data, analyze the data and generate learning analytics for delivery to students and their instructors or researchers on demand or when conditions are matched.

Stephanie Teasley in *Student Facing Dashboards: One Size Fits All?* reports that early implementations of dashboards provide mixed results about the effects of their use. In particular, the 'one-size-fits-all' design of many existing systems is questioned based on research on performance feedback and student motivation, which has shown that both internal and external student-level factors affect the impact of feedback interventions, especially those using social comparisons. She asserts that integrating data from student information systems into underlying algorithms to produce personalized dashboards may mediate the possible negative effects of feedback, especially comparative feedback, and support more consistent benefits from the use of such systems.

5 Integrative Review

Acknowledging that various disciplines attempt to infer learning from big data using different methodologies, the next authors provide a framework for moving transdisciplinary conversations forward in research collaborations. In their integrative review entitled *Inferring learning from big data: The importance of a transdisciplinary and multidimensional approach* Jason M. Lodge, Sakinah S. J. Alhadad, Melinda J. Lewis and

Dragan Gašević discuss the need for systematic collaboration across different paradigms and disciplinary backgrounds in interpreting big data for enhancing learning.

I trust that you will find one or more of these projects and reports to be of interest and will lead you to follow these researchers as data science evolves in higher education learning and teaching.

David Gibson

References

Baker, R., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining, 1,* 3–17. doi:10.1109/ASE.2003.1240314.

Buckingham Shum, S., Hawksey, M., Baker, R. S., Jeffery, N., Behrens, J. T., & Pea, R. (2013). Educational data scientists: A scarce breed. In *3rd international conference on learning analytics and knowledge, LAK 2013* (pp. 278–281). doi:10.1145/2460296.2460355.

Gibson, D. (2012). Game changers for transforming learning environments. In F. Miller (Ed.), *Transforming learning environments: Strategies to shape the next generation (advances in educational administration)* (Vol. 16, pp. 215–235). Bingley: Emerald Group Publishing Ltd. doi:10.1108/S1479-3660(2012)0000016014.

Gibson, D., & Ifenthaler, D. (2017). Preparing the next generation of education researchers for big data in higher education. In B. Kei Daniel (Ed.), *Big data and learning analytics: Current theory and practice in higher education* (pp. 29–42). Berlin: Springer.

Gibson, D., & Webb, M. (2015). Data science in educational assessment. *Education and Information Technologies, 24*(4), 697–713. doi:10.1007/s10639-015-9411-7.