Chapter 27: A Critical Perspective on Learning Analytics and Educational Data Mining

Rita Kop[1], Helene Fournier[2], Guillaume Durand[2]

[1]Faculty of Education, Yorkville University, United Kingdom
[2]Information and Communications Technologies, National Research Council of Canada, Canada

ABSTRACT

In our last paper on educational data mining (EDM) and learning analytics (LA; Fournier, Kop & Durand, 2014), we concluded that publications about the usefulness of quantitative and qualitative analysis tools were not yet available and that further research would be helpful to clarify if they might help learners on their self-directed learning journey. Some of these publications have now materialized; however, replicating some of the research described met with disappointing results. In this chapter, we take a critical stance on the validity of EDM and LA for measuring and claiming results in educational and learning settings. We will also report on how EDM might be used to show the fallacies of empirical models of learning. Other dimensions that will be explored are the human factors in learning and their relation to EDM and LA, and the ethics of using "Big Data" in research in open learning environments.

Keywords: Educational data mining (EDM), big data, algorithms, serendipity

The past ten years have been interesting in the fields of education and learning technology, which seem to be in flux. Whereas past research in education related to the educational triangle of learner, instructor, and course content (Kansanen & Meri, 1999; Meyer & Land, 2006), newly developed technologies put an emphasis on other dimensions influencing learning; for instance, the learning context or learning setting and the technologies being used (Bouchard, 2013). Fenwick (2015a) posits that humans and the technologies they use are not separate entities: "material and social forces are interpenetrated in ways that have important implications for how we might examine their mutual constitution in educational processes and events" (p. 14). Not only is there an interaction between humans and materials such as technology but also a symbiotic relationship.

New technologies have moved us from an era of scarcity of information to an era of abundance (Weller, 2011). Social media now make it possible to communicate across networks on a global scale, outside the traditional classroom bound by brick walls. Communication on such a global scale would have been unimaginable not long ago. Data and data storage have evolved under the influence of emerging technologies. Instead of capturing data and storing it in a database, we now deal with large data-streams stored in the cloud, which might be represented and visualized using algorithms and machine learning. This presents interesting opportunities to learn from data, revealing with it hidden insights but important challenges as well.

Questions have been raised about how stakeholders – learners, educators, and administrators – in the educational process might manage and access all these levels of information and communication effectively. Computer scientists have suggested opportunities for automated data filtering and analysis that could do exactly that: sift through all data available and provide learners with connections to and recommendations for their preferred information, people, and tools, and in doing so personalize the learning experience and aid learners in the management and deepening of their learning (Siemens, Dawson, & Lynch, 2013). In addition, examples of research using huge institutional

datasets are emerging, made possible by accessing data from traces left behind by learner activity (Xu & Smith Jaggars, 2013).

In discussing changes big data might force onto professional practice, Fenwick (2015b) highlights, for instance, the "reduction of knowledge in terms of decision making. Data analytics software works from simplistic premises: that problems are technical, comprised of knowable, measurable parameters, and can be solved through technical calculation. Complexities of ethics and values, ambiguities and tensions, culture and politics and even the context in which data is collected are not accounted for" (p. 70).

This is an important issue. She further emphasizes that the current developments involving data might change our everyday practice in ways that may not quite be understood when implemented. For instance, she highlights equality issues that arise when there is a dependence on comparison and prediction (Fenwick, 2015b). Moreover, her research led to the conclusion that research methodologies taught to prospective educators and educational researchers are completely inadequate in dealing with the big datasets available to enhance their practice. Furthermore, she wonders about "the level of professional agency and accountability. Much data accumulation and calculation is automated, which opens up new questions about the autonomy of algorithms and the attribution of responsibility when bad things happen" (p. 71). These are serious questions that need careful consideration. This chapter will address some of the challenges related to educational data mining and analytics using large datasets in research and user data in algorithms for learner support. It will also explore the impact of automation and the possible dehumanizing effects of replacing human communication and engagement in learning with technology.

THE OPPORTUNITIES OF EDUCATIONAL DATA MINING IN LEARNING MANAGEMENT

Reliability and Validity

Educational data mining (EDM) is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students and the settings in which they learn (Ed Tech Review, 2016). Educational data mining is wider than its name would imply and it goes beyond the scope of simply mining educational data for information retrieval and building a better understanding of learning mechanisms. Hence, EDM also aims at developing methods and models to predict learner behaviours

using machine learning and statistical approaches governed by scientific concerns related to validity, reproducibility, and generalizability.

Learning analytics (LA) is closely related to the field of EDM and is concerned with 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 (Long & Siemens, 2011). EDM techniques and LA are being used to augment the learning process. These seem promising in aiding in the provision of effective student support, and although there is promise that these new developments might enhance education and learning, major challenges have also been identified.

To some extent, EDM is not only a research area, thriving due to the prolific contributions of researchers from all around the world, but also a science. Recently, Britain's Science Council defined science as "the pursuit of knowledge and understanding of the natural and social world following a systematic methodology based on evidence" (British Science Council, 2009). Evidence is a requirement to any claim made in the field; as in any other scientific domain, educational data mining and analytics researchers require evidence to support or reject claims and discoveries drawn from or validated by educational data.

However, a common definition of what makes good or poor evidence is not that obvious in the EDM and LA research community, which has brought together scientists from "hard" (Computer Science) and "soft" sciences (Education). We will provide here some examples of inconsistencies and procedural flaws that we have come across during our own research. Thanks to data sharing, Long and Aleven (2014) were able to contradict learning claims of a gamified approach in an intelligent tutoring system. However, sometimes sharing datasets is not sufficient; some research work requires extensive preprocessing as several choices (biases) are made during those steps that may be hard to define clearly in a research paper. Another team trying to prepare the dataset following the same rules, therefore, might not manage to do so. The nature of the software used in preprocessing can also have an impact. The implementation of key methods can vary when using R, SPSS, Matlab, and other tools, leading to potentially different conclusions.

Another contentious aspect is the *a priori* assumption or "ground truth" that a statistical model would be built upon. This is the case, for example, in competency frameworks where mapping between items and skills defined by human experts is questionable (Durand, Belacel, & Goutte, 2015). Skills, like many other latent traits, are sometimes hard to characterize. To that end, PSLC Data Shop offers an incredible environment

for learning experts to test their competency frameworks, as the observed results obtained by students help them to improve and share their mapping. It is also a great tool to share datasets among EDM and LA practitioners since sharing datasets is valuable in identifying problems. Sharing should become commonplace and no major published results should be seriously considered without the possibility for other teams to validate the claims.

Some other issues might raise questions, such as considering statistical studies and particularly linear correlational measures. EDM gathers researchers with different practices and with different perspectives on what could or should not be used as evidence. While in "hard science" a significant Pearson correlation below r=.5 would be systematically considered weak, it is usual in "soft science" to consider r=.3 values to be strong, especially regarding personality traits. Psychologists even call this .3 threshold, the "personality coefficient" because most relationships between personality traits and behaviours tend to be around that value, including the relationship between competency and performance (Mischel, 1968, p. 78). Work done in EDM regarding sentiment analysis (Wen, Yang, & Rosé, 2014) provides such an example, where it is difficult to provide computational outcomes from meaningful "soft" science research results on dropout rates in MOOCs. It might also be suggested that the topic under investigation would be better researched through qualitative techniques. However, relationships in this form of quantitative EDM remain weak when the intent is to infer predictions. Specifically, a .3 correlation that by definition explains 9% of the variance in the criterion may be of limited value in predictions in the area of sentiment analysis.

Several statistics tests can be significant as well but not really truthful regarding the accuracy of the results. El Emam (1998) evaluated how the Chi-Square test could be misleading in evaluating the predictive validity of a classifier, showing that same Chi-Square results could prove either strong or weak accuracies. More recently, Gonzalez-Brenes and Huang (2015) proposed the Leopard metric as a standard way of evaluating adaptive tutoring systems and increasing the evaluation results of the predictive accuracy of the system by evaluating their usefulness. They proposed to evaluate the amount of effort required by learners in those systems to achieve learning outcomes. After all, usefulness measures might be what the people using the systems are most interested in.

To that end, Ryan Baker, one of the most prominent researchers of the EDM community, in his MOOC entitled "Big Data in Education" provides examples and good practical advice to help researchers understand and check more wisely the validity of their models (generalizability, ecological, construct, and predictive, substantive, and content validity). In his course, Baker emphasized using Kappa and even better A' to measure respectively how a "detector is better than chance" and "the probability a detector will correctly identify" a specific trait to overcome some of the flaws of other metrics for classifiers like accuracy, ROC, precision, and recall or Chi-Square (Ocumpaugh, Baker, Gowda, Heffernan, & Heffernan, 2014, p. 492). However, A' and Kappa use seems limited in EDM publications so far.

Critically, we would like to emphasize the importance of research integrity. It might be appealing in EDM and LA to provide "made up" results. Our own work has shown that obtaining tangible results usually requires many attempts, much work is done without any guaranty of success, and the validation process appears problematic. So far, no major cases of falsifying results have been revealed but providing guidelines regarding transparency should be of greater concern to avoid potential future cases of fraud and misconduct, as observed in other scientific areas (Gupta, 2013).

We argue that in the developing fields of LA and EDM, the scientific ideal is ambitious and requires that researchers carefully check the scientific robustness of their claims. Even though research is fuelled by grants based on promises that it will impact human learning in the near future, it is important to take the time to safeguard the scientific integrity of the emerging fields of LA and EDM. This requires careful consideration by us all of the methods used and the results obtained.

The Challenges of Qualitative Data Analysis

If we look at the development of educational research over the past decades, there is a distinct movement from quantitative towards qualitative research (Gergen, Josselson, & Freeman, 2015). Psychologists increasingly support the idea that the intricacies of learning and knowing cannot be determined by testing individuals' behaviour alone. The study of the richness of learner actions and thinking in relation to the society they live in, and their communications with those in their knowledge networks, provides a much deeper, more inclusive, and critically cultural understanding of people's knowledge development and learning (Christopher, Wendt, Marecek, & Goodman, 2014; Denzin & Lincoln, 2011; Gergen et al., 2015).

The current technology-rich learning environment is not just a walled-in classroom but involves global network communications too; these encompass reflexive narrative and rich imagery, challenging researchers to re-invent their research methodologies. Moving beyond end-of-course surveys to reveal student satisfaction, mining the data produced by learners, analyzing the

narrative, images, and visualizations produced during the online learning experience – all of these offer options for understanding the rich tapestry of learning interactions. Analyzing the fundamental dimensions in the changing assemblages of words and images on social media that now form part of the learning environment might get more to the heart of the learning process than official course evaluations.

Research on PLENK2010 and CLOM REL 2014, two massive open online courses (MOOCs), has highlighted the challenges that such research involves (Fournier & Kop, 2015; Kop, Fournier, & Durand, 2014). Previous MOOC research provided both bigger and richer datasets than ever before, with powerful tools to visualize patterns in the data, especially on digital social networks. The work of uncovering such patterns, however, provided more questions than answers from the pedagogical and technical contexts in which the data were generated. Moving towards a qualitative approach in trying to understand why MOOC participants produced the data that they did prompted a critical reflection on what big data, EDM, and LA could and could *not* tell us about complex learning processes and experiences.

Boyd (2010) expressed it in the following way:

> Much of the enthusiasm surrounding Big Data stems from the opportunity of having easy access to massive amounts of data with the click of a finger. Or, in Vint Cerf's words, "We never, ever in the history of mankind have had access to so much information so quickly and so easily." Unfortunately, what gets lost in this excitement is a critical analysis of what this data is and what it means. (p. 2)

In dealing with so much data and information so quickly, researchers need to envisage the optimal processes and techniques for translating data into understandable, consumable, or actionable modes of representation in order for results to be useful and accessible for audiences to digest. The ability to communicate complex ideas effectively is critical in producing something of value that translates research findings into practice. Questions have been raised about how stakeholders in the educational process (i.e., learners, educators, and administrators) might access, manage, and make sense of all these levels of information effectively; EDM and LA methods hint at how automated data filtering and analysis could do exactly that. This can lead to potentially rich inferences about learning and learners but also raise many new interesting research questions and challenges in the process. In so doing, researchers must strive to demonstrate how the data are meaningful, as well as appealing to various stakeholders in the educational process while engaging in responsible innovation with

thoughtful research designs and implementations (Berland, Baker, & Blikstein, 2014).

Algorithms, Serendipity, and the "Human" in Learning: A Critical Look at Learning Analytics

A body of literature is slowly developing in EDM and LA. Essentially, it is not easy to use technology to analyze learning or use predictive analytics to advance learning. Issues around the development of algorithms and other data-driven systems in education lead to questions about what these systems actually replace and whether this replacement is positive or negative. Secondly, who influences the content of data-driven systems and what value might they add to the educational process?

In online education, but also in a connectivist networked environment (Jones, Dirckinck-Holmfeld, & Lindström, 2006), communication and dialogue between participants in the learning endeavor have been at the heart of a quality learning experience. This human touch is a necessary component in developing learning systems and environments (Bates, 2014). The presence and engagement of knowledgeable others has always been seen as vital to extend the ideas, creativity, and thinking of participants in formal learning settings, but also in online networks of interest (Jones et al., 2006).

When developing data-driven technologies for learning, it seems important to harness this human element somehow for the good of the learning process. This means that in the filtering of information, or the asking of Socratic questions, the aggregation of information should be mediated via human beings (Kop, 2012). Social microblogging sites such as Twitter have been shown to do this successfully, as "followers," who provide information and links to resources, have been chosen by the user and are seen to be valuable and credible (Bista, 2014; Kop, 2012; Stewart, 2015). In algorithms, these judgements are difficult to achieve, but perhaps a combination of recommender systems, based on data, and support and scaffolding applications based on communication, would facilitate this.

It is important to consider who influences the content of data-driven systems and what value they might add to the educational process. Furthermore, not only do the affordances and effectiveness of new technologies need to be considered, but also a reflection on the ethics of moving from a learning environment characterized by human communication to an environment that includes technical elements over which the learner has little or no control.

One of the problems highlighted in the development of algorithms is the introduction of researcher biases in the tool, which could affect the quality of the recommendation or search result (Hardt, 2014). Proper

training of the people working with the data can make all the difference (Fenwick, 2015b; Boyd & Crawford, 2012). Presently, computer scientists and mathematicians, who do not necessarily have a background in the social sciences, produce the applications. As Boyd and Crawford (2012) compellingly argue:

> When computational skills are positioned as the most valuable, questions emerge over who is advantaged and who is disadvantaged in such a context. This, in its own way, sets up new hierarchies around "who can read the numbers," rather than recognizing that computer scientists and social scientists both have valuable perspectives to offer. Significantly, this is also a gendered division. Most researchers who have computational skills at the present moment are male and, as feminist historians and philosophers of science have demonstrated, who is asking the questions determines which questions are asked. (p. 674)

Boyd and Crawford (2012) suggest that computer scientists and social scientists should work together to develop bias-free, high quality analytics tools, and that teamwork with people in different fields might also be fruitful for the mining and analysis of big data. Of course, the expansion and availability of data has also made it attractive to make use of them, but there are again some challenges. Human beings for the most part get their information from sources that they trust, but as Fenwick (2015b) suggests, the use of new techniques might change "everyday practice and responsibilities in ways that may not be fully recognised" (p. 71). She highlights, for instance, that a reliance on comparison and prediction "can be self-reinforcing and reproductive, augmenting path dependency and entrenching existing inequities," especially if the people producing the algorithms are not aware of the reinforcement of stereotypes when big data is not used carefully.

Furthermore, we should not underestimate the fact that most of the algorithms currently in use were produced for economic gain and not to enhance deeper levels of learning or add value to society. As argued by Kitchin (2015), "Software is not simply lines of code that perform a set of instructions, but rather needs to be understood as a social product that emerges in contingent, relational and contextual ways, the outcome of many minds situated with diverse social, political and economic relations" (p. 5). Clearly, the development of automated algorithm systems has another inherent problem wherein it might be hard to point a finger towards who is responsible when things go wrong.

Some Ethical Considerations

Open learning environments combined with powerful data analysis tools and methods bring new affordances and support for learning but also highlight important ethical issues and challenges that move learners from an environment characterized by human communication to one that includes technical elements over which the learner has little or no control. Much of the commercial effort in Web development is informed by big data and is lacking in any innovative educational insights (Atkinson, 2015). We agree that "It is the scholarship and research informed learning design itself, grounded in meaningful pedagogical and andragogical theories of learning that will ensure that technology solutions deliver significant and sustainable benefits" to education (Atkinson, 2015, p. 7).

The dynamic pace of technological innovation, including EDM and LA, also requires the safeguarding of privacy in a proactive manner. In order to achieve this goal, researchers and system designers in the fields of EDM and advanced analytics must practice responsible innovation that integrates privacy-enhancing technologies directly into their products and processes (Cavoukian & Jonas, 2012). According to Oblinger (2012), "Analytics is a matter of culture – a culture of inquiry: asking questions, looking for supporting data, being honest about strengths and weaknesses that the data reveals, creating solutions, and then adapting as the results of those efforts come to fruition" (p. 98).

With this in mind, we strongly recommend that those designing and building next generation analytics ensure that they are informed by Privacy by Design. This entails mindfulness and responsible practice involving accountability, research integrity, data protection, privacy, and consent (Cavoukian & Jonas, 2012; Cormack, 2015). The line between private and public data is increasingly becoming blurred as more opportunities to participate in open learning environments are created and as data about participants, their activities, their interactions, and their behaviours are made accessible through social media, such as Facebook, Twitter, Google, and potentially any other social media tool available online. In the context of big data, we agree with the European Data Protection Supervisor (2015) who states that "People want to understand how algorithms can create correlations and assumptions about them, and how their combined personal information can turn into intrusive predications about their behaviour" (p. 10).

CONCLUSION

Significant questions about truth, control, transparency, and power in big data studies also need to be addressed. Pardo and Siemens (2014) maintain that keeping too much data (including student digital data, privacy-sensitive data) for too long may actually be harmful and lead to mistrust of the system or institution that has been entrusted to protect personal data. Discussions around big data ethics have underscored important methodological concerns related to data cleaning, data selection and interpretation (Boyd & Crawford, 2012), the invasive potential of data analytics, as well as the potential dehumanizing effects of replacing human communication and engagement with automated machine-learning algorithms and feedback. Researchers and developers must be mindful of the affordances and limitations of big data (including data mining and predictive learning analytics) in order to construct useful future directions (Fenwick, 2015b). Researchers should also work together in teams to avoid some of the inherent fallacies and biases in their work, and to tackle the important issues and challenges in big data and data-driven systems in order to add value to the educational process.

REFERENCES

Atkinson, S. P. (2015). Adaptive learning and learning analytics: A new learning design paradigm. BPP Working paper. https://spatkinson.files.wordpress.com/2015/05/atkinson-adaptive-learning-and-learning-analytics.pdf

Bates, T. (2014). Two design models for online collaborative learning: Same or different? http://www.tony-bates.ca/2014/11/28/two-design-models-for-online-collaborative-learning-same-or-different/

Berland, M., Baker, R. S., & Blikstein, P. (2014). Educational data mining and learning analytics: Applications to constructionist research. *Technology, Knowledge and Learning*, 19, 206–220.

Bista, K. (2014). Is Twitter an effective pedagogical tool in higher education? Perspectives of education graduate students. *Journal of the Scholarship of Teaching and Learning*, 15(2), 83–102. http://files.eric.ed.gov/fulltext/EJ1059422.pdf

Bouchard, P. (2013). Education without a distance: Networked learning. In T. Nesbit, S. M. Brigham, & N. Taber (Eds.), *Building on critical traditions: Adult education and learning in Canada.* Toronto: Thompson Educational Publishing.

Boyd, D. (2010). Privacy and publicity in the context of big data. Paper presented at the 19th International Conference on World Wide Web (WWW2010), 29 April 2010, Raleigh, North Carolina, USA. http://www.danah.org/papers/talks/2010/WWW2010.html

Boyd, D., & Crawford, K. (2012). Critical equations for Big Data. *Information, Communication & Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878

British Science Council. (2009). What is science? http://www.sciencecouncil.org/definition

Cavoukian, A., & Jonas, J. (2012). Privacy by design in the age of big data. https://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf

Cormack, A. (2015). A data protection framework for learning analytics. Community.jisc.ac.uk. http://bit.ly/1OdIIKZ

Christopher, J. C., Wendt, D. C., Marecek, J., & Goodman, D. M. (2014) Critical cultural awareness: Contributions to a globalizing psychology. *American Psychologist, 69*, 645–655. http://dx.doi.org/10.1037/a0036851

Denzin, N., & Lincoln, Y. (Eds.) (2011). The Sage handbook of qualitative research (4th ed.). Thousand Oaks, CA: Sage.

Durand, G., Belacel, N., & Goutte, C. (2015) Evaluation of expert-based Q-matrices predictive quality in matrix factorization models. *Proceedings of the 10th European Conference on Technology Enhanced Learning* (EC-TEL '15), 15–17 September 2015, Toledo, Spain (pp. 56–69). Springer. doi:10.1007%2F978-3-319-24258-3_5

Ed Tech Review (2016). Educational Data Mining (EDM). http://edtechreview.in/dictionary/394-what-is-educational-data-mining

El Emam, K. (1998). The predictive validity criterion for evaluating binary classifiers. *Proceedings of the 5th International Software Metrics Symposium* (Metrics 1998), 20–21 November 1998, Bethesda, MD, USA (pp. 235–244). IEEE Computer Society. http://ieeexplore.ieee.org/document/731250/

European Data Protection Supervisor. (2015). Leading by example: The EDPS strategy 2015–2019. https://secure.edps.europa.eu/EDPSWEB/edps/site/mySite/Strategy2015

Fenwick, T. (2015a). Things that matter in education. In B. Williamson (Ed.), *Coding/learning, software and digital data in education*. University of Stirling, UK. http://bit.ly/1NdHVbw

Fenwick, T. (2015b). Professional responsibility in a future of data analytics. In B. Williamson (Ed.) *Coding/learning, software and digital data in education*. University of Stirling, UK. http://bit.ly/1NdHVbw

Fournier, H., & Kop, R. (2015). MOOC learning experience design: Issues and challenges. *International Journal on E-Learning*, 14(3), 289–304.

Gergen, J. K., Josselson, R., & Freeman, M. (2015). The promise of qualitative inquiry. *American Psychologist*, 70(1), 1–9.

Gonzalez-Brenes, J., & Huang, Y. (2015). Your model is predictive – but is it useful? Theoretical and empirical considerations of a new paradigm for adaptive tutoring evaluation. In O. C. Santos, J. G. Boticario, C. Romero, M. Pechenizkiy, A. Merceron, P. Mitros, J. M. Luna, C. Mihaescu, P. Moreno, A. Hershkovitz, S. Ventura, & M. Desmarais (Eds.), *Proceedings of the 8th International Conference on Education Data Mining* (EDM2015), 26–29 June 2015, Madrid, Spain (pp. 187–194). International Educational Data Mining Society.

Gupta, A. (2013). Fraud and misconduct in clinical research: A concern. *Perspectives in Clinical Research*. 4(2), 144–147. doi:10.4103/2229-3485.111800.

Hardt, M. (2014). How big data is unfair: Understanding sources of unfairness in data driven decision making. https://medium.com/@mrtz/how-big-data-is-unfair-9aa544d739de

Jones, C., Dirckinck-Holmfeld, L., & Lindström, B. (2006). A relational, indirect, meso-level approach to CSCL design in the next decade. *International Journal of Computer Supported Collaborative Learning*, 1(1), 35–56.

Kansanen, P., & Meri, M. (1999). The didactic relation in the teaching–studying–learning process. *TNTEE Publications*, 2, 107–116. http://www.helsinki.fi/~pkansane/Kansanen_Meri.pdf

Kitchin, R. (2015). Foreword: Education in code/space. In B. Williamson (Ed.), *Coding/learning, software and digital data in education*. University of Stirling, UK. http://bit.ly/1NdHVbw

Kop, R. (2012). The unexpected connection: Serendipity and human mediation in networked learning. *Educational Technology & Society*, 15(2), 2–11.

Kop, R., Fournier, H., & Durand, G. (2014). Challenges to research in massive open online courses. *Merlot Journal of Online Learning and Teaching*, 10(1). http://jolt.merlot.org/vol10no1/fournier_0314.pdf

Long, Y., & Aleven, V. (2014). Gamification of joint student/system control over problem selection in a linear equation tutor. In S. Trausan-Matu, K. E. Boyer, M. Crosby, & K. Panourgia (Eds.), *Proceedings of the 12th International Conference on Intelligent Tutoring Systems* (ITS 2014), 5–9 June 2014, Honolulu, HI, USA (pp. 378–387). New York: Springer. doi:10.1007/978-3-319-07221-0_47

Long, C., & Siemens, G. (2011, September 12). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, 46(5). http://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education

Meyer, J. H. F., & Land, R. (2006). *Overcoming barriers to student understanding: Threshold concepts and troublesome knowledge*. London: Routledge.

Mischel, W. (1968). *Personality and assessment*. London: Wiley.

Oblinger, D. G. (2012, November 1). Analytics: What we're hearing. *EDUCAUSE Review*. http://er.educause.edu/articles/2012/11/analytics-what-were-hearing

Ocumpaugh, J., Baker, R., Gowda, S., Heffernan, N., & Heffernan, C. (2014). Population validity for educational data mining models: A case study in affect detection. *British Journal of Educational Technology*, 45(3), 487–501.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.

Siemens, G., Dawson, D., & Lynch, L. (2013). *Improving the quality and productivity of the higher education sector: Policy and strategy for systems-level deployment of learning analytics*. SoLAR. https://www.itl.usyd.edu.au/projects/SoLAR_Report_2014.pdf

Stewart, B. (2015). Open to influence: What counts as academic influence in scholarly networked Twitter participation. *Learning, Media and Technology*, 40(3), 287–309. doi:10.1080/17439884.2015.1015547

Weller, M. (2011). A pedagogy of abundance. *Spanish Journal of Pedagogy*, 249, 223–236.

Wen, M., Yang, D., & Rosé, C. (2014). Sentiment analysis in MOOC discussion forums: What does it tell us? In J. Stamper, Z. Pardos, M. Mavrikis, & B. M. McLaren (Eds.), *Proceedings of the 7th International Conference on Educational Data Mining* (EDM2014), 4–7 July, London, UK (pp. 130–137). International Educational Data Mining Society.

Xu, D., & Smith Jaggars, S. (2013). Adaptability to online learning: Differences across types of students and academic subject areas. CCRC Working Paper No. 54. Community College Research Center, Teachers College, Columbia University. http://anitacrawley.net/Reports/adaptability-to-online-learning.pdf