

Framing Learning Analytics and Educational Data Mining for Teaching: Critical Inferencing, Domain Knowledge, and Pedagogy



Owen G. McGrath

This chapter reviews key challenges of learning analytics and educational data mining. It highlights early generation learning analytics pitfalls that could compromise the future of their use in technology-delivered instruction, especially if teachers are not well trained and adequately equipped with both technical and sociocritical literacy of this new field. Among the issues are potential for bias and inaccuracy in the algorithms involved, the propensity toward closed proprietary systems whose algorithms cannot be scrutinized, and the paucity of learning models typically considered. The new learning analytics and educational data mining systems bring with them a set of claims, aspirations, and mystique. These underlying technologies could be harbingers of future breakthroughs: a new generation of artificial intelligence systems adaptively responding to students' interactions with online teaching environments. However, current system implementations and research studies reveal an immaturity of methods and a tendency to focus narrowly on a small range of easily tracked user behaviors that are only indirectly associated with learning (Blikstein, 2013). The initial wave of studies and proof-of-concept systems seem at times like technologies in search of a problem, i.e., hammers in search of nails. There is a familiar risk here in allowing the technology developers to set the agenda—a risk that doomed previous generations of intelligent learning systems. Unless domain experts and stakeholders (i.e., teachers and teacher researchers) are trained up to critique and shape the design of these new technologies, the resulting systems will likely repeat the failures of the past: deployed as black box “expert systems” that confuse, constrain, or supplant teachers altogether while also jeopardizing the privacy and agency of students. Instead, this chapter argues, these technologies need to be conceived and designed within a broader context of supporting teaching, particularly teacher decision-making. The promise these systems hold can only be realized

O. G. McGrath (✉)

Educational Technology Services, UC Berkeley, Berkeley, CA, USA
e-mail: omcgrath@berkeley.edu

© Springer Nature Singapore Pte Ltd. 2018
J. M. Spector et al. (eds.), *Frontiers of Cyberlearning*, Lecture Notes in Educational Technology, https://doi.org/10.1007/978-981-13-0650-1_2

if they are designed for the domain experts, i.e., teachers. Teacher training programs, in turn, need to add data science to the curriculum.

By many accounts, a measurement revolution is taking place in global secondary and tertiary education (Long & Siemens, 2011; Daniel, 2015). As with earlier technology movements in education, the actual outcome will depend on how well teachers are trained to understand, influence, and make use of the technology (Cuban, 2001; Jung, 2005; Kenny, 2006; Kozma, 2008). The risks around decision-making systems-based big data algorithms are only just starting to be understood, despite their widespread use in many industries (Jagadish, 2015; O'Neil, 2016). Given the black box approach to commercial learning analytics systems, the algorithms and models used for sorting and labeling students could also go by uninspected like a proprietary secret sauce. It would be all too easy for educators to stand aside as unquestioning and passive end users of these opaque systems deployed at the institutional level. Instead, this chapter argues, learning analytics and educational data mining systems should be designed for and by the domain experts, i.e., teachers and they should be implemented with transparency and openness so that their algorithms can be scrutinized and tested for fairness. If designed from the start to support the actual needs of teachers, these systems could be engineered to support teachers' inquiry and decision-making in pursuit of instructional effectiveness (Kumar et al., 2015). With data-driven information technologies as the key enablers, the learning analytics and educational data mining movement could offer new ways of asking questions about what gets taught and learned in school settings.

This technology-supported inquiry could yield a new understanding of learning outcomes, teacher effectiveness, personalization of learning, and perhaps even the core assumptions of requiring in-person education, which has been an organizing principle for most institutions of higher learning for centuries (Brown & Kurzweil, 2017). But how to engineer such systems reliably, what to measure, and how best to support teaching? These key questions are only just starting to be asked at this early stage. Early implementations and research studies of learning analytics and educational data mining reveal a narrowness of methods and a tendency to focus narrowly on a small range of easily tracked user behaviors such as the number of times they logged in, visited a web page, or lingered on a display of information. These limits have constrained the context and variables considered, and even so far, the new field has tended toward peripheral (albeit measurable) variables and narrow range of models of teaching and learning.

This chapter makes a case for teacher training and teacher research programs to engage with learning analytics and educational data mining not only to be critically aware of the key challenges revealed in early generation efforts but also to help shape the future of these new technologies. First is the need for teachers, as social scientists, to be critically literate in terms of the new technology: the role of algorithms, the means of inferencing, and the methods training with data. Learning analytics and educational data mining certainly bring with them a set of claims, aspirations, and mystique. Teachers in particular and teacher researchers as well need to consider critically key questions of fairness, reliability, and validity that lurk within these technologies. Critical literacy of these technologies, when used in sup-

porting decision-making around instructional attainment and effectiveness, must be built upon familiar fundamental concerns with bias, model selection, validity, and reliability. In particular, teachers should feel empowered to consider critically the quality and provenance of the massive data used in these systems, the models of successful and failed learning used, the rate and accumulation of error, etc. Moreover, professional educators—as data scientists—also need to be empowered to call for the ongoing audit and scrutiny of the algorithms and data models employed. A second key area discussed in this chapter involves the theoretical models of teaching and learning upon which these new decision-making support systems are built. To date, most of the systems in this new era of learning analytics and educational data mining system are limited by their sources and methods to dealing only with a narrow range of directly observable online actions of learners, their outward digital behaviors, and some institutionally recorded categorical attributes. The data for these online behavioral traces are typically then analyzed in terms of correlates with assessment data, academic achievement measures, or normative digital behaviors of “successful” students.

Largely missing from the current focus on students’ recordable interactions with online systems is much in the way of significant theorizing or even informed speculation about the relationship of teaching to student behavior in the broader contexts (e.g., classroom, institution), teacher attributes, or the material being taught. Teaching strategies, interactions, decision-making, attributes, etc., are absent as data or variables. Instead, the typical educational contexts considered are limited in scope to traces of online interactions, formative or summative assessment measures, and institutionally held categorical data (e.g., grade level, gender, SES, standardized achievement score history). The resulting approach toward teaching and learning that are implied in most early generation learning analytics and educational data mining systems is a simplistic, teacher-free view of learning as incremental behavioral pathways online that are either rewarded or remediated based on norms formed and update along the way through correlates of online success.

As advocated in this chapter, an alternative and more promising approach for the future envisions the scope and design of these learning analytics and educational data mining systems framed more broadly around questions and variables that more relevant to practitioners in the domain. These would include areas in which teachers bring together their content and pedagogical knowledge to design and carry out instructional activities: e.g., their structuring of material, selection of media and sequencing, the teacher/student discourse patterns, etc. The fields of teacher research and teacher training can bring to bear the domain knowledge to provide a crucial research and advocacy role that promotes and advances attention to such models, as opposed to the rather limited pedagogy (e.g., online lectures interspersed with computer-marked assignments) focused upon thus far (Daniel, 2010, 2012). What is needed are teaching and learning paradigms of content knowledge such that the design and use of the system will be based on a framework that considers both content knowledge and pedagogy (Shulman, 1986; Carlsen, 2001; Kleickmann et al., 2013). Priority should go to teachers’ reflection and decision-making with helpful insights

into the relationship between their understanding of subject matter and the instruction they provide to students.

As we will see in Section 1, to take up such an agenda would be a timely move for the fields of teacher training and research on teaching, given the rise of technology-augmented instruction in all levels of schooling. Indeed, as Section 2 will show, the convergence of networked information technology underlying learning analytics and educational data mining offers significant opportunities for expansive improvements to teaching and learning whether in traditional or virtual schools. Section 3 points particularly to the need for developing in teachers and teacher researchers the ability to consider critically both how these systems work and how educational data is mined. Section 4 looks at some guiding principles for the teacher training and teacher research fields' appropriate roles in the learning analytics and educational data mining era. These principles, framed in terms of teaching and learning, require paying attention to both when turning data into knowledge useful for decision-making.

1 Wired and Virtual Schools

Underlying and enabling the rise of learning analytics and educational data mining are the networked information technologies now reaching into formal education worldwide. In North America, secondary and tertiary education teaching and learning activities are increasingly carried out through and supported by Information and Communication Technology (ICT) both inside and outside the classroom. A convergence of enabling technologies (e.g., the Internet, mobile phones, tablet computing devices, cloud computing, satellite-based Internet access) has opened up transcendent possibilities for using networked computing and communications technologies to extend teaching and learning opportunities in unprecedented ways. In particular, the coming decades of ICT for education will likely be remembered as the dawn of technology-augmented teaching and fully online instruction. Accredited secondary and tertiary school systems delivering and managing instruction via technology within the classroom and blended or fully online instruction outside is becoming commonplace. With the rise of cyber-infrastructure in secondary and tertiary education, new opportunities surface when it comes to understanding learner's online activities. How, where, and when learner activity is captured and analyzed in academic online systems is particularly critical in these networked systems. On the flipside, the flexibility that Internet-based systems allows for in promoting easy integration of different technologies and platforms has repercussions for the engineering of these new systems: around the clock access to a multitude of distributed users can result in huge volumes of online learning data.

Whether it is online learning in traditional schools virtual schools or mega-schools, ICT-based online teaching and learning offer compelling opportunities to consider new approaches to teaching and learning, as a diverse and growing group of educational leaders and analysts agree (Moe & Chubb 2009; Daniel 2010). Whole books could be written in describing the many key enabling technologies that are allowing

for online learning: the Internet, mobile phones, tablet computing devices, cloud computing, satellite-based Internet, etc. Whole books have been written about the wide range of possible teaching and learning modes, methods, and models that online teaching and learning might use. Vigorous debates and wide-ranging proposals already abound for possible organizational structures, methods of delivery, modes of institutional alignment, and assessment models for best implementing ICT-based online schools and ICT-based teacher training for secondary and tertiary education (Bramble & Panda, 2008). Across many of these varied proposals is also a shared sense that the sophistication and reach of ICT creates a historic opportunity to focus on designing personalized learning environments with revolutionary support for teacher decision-making.

2 Learning Analytics and Educational Data Mining

With the spread of networked information technology into secondary and tertiary education, the fields learning analytics and educational data mining emerged in the late 2000s as subfields of a wider movement toward web analytics and online usage data mining (Bach, 2010). It would be difficult to overstate the importance that web usage analytics and data mining already have as constitutive components of today's web-based e-commerce models and social computing paradigms. Tremendous amounts of money and research are being directed toward the art and practice of probing deeply into the mountains of activity data users leave behind in visiting online material. More controversial is the increasing deployment of browsing analytics and data mining for surveillance and profiling of users. Debates about the pros and cons of these kinds of tracking and monitoring technologies are only just beginning.

Although they are subfields of web usage analytics in general, learning analytics and educational data mining are not the same; it should be conceded. Nevertheless, they are paired throughout this chapter mainly because of the shared set of issues and challenges they present in their common implementations so far. The terms learning analytics and educational data mining have come to refer generally to a set of somewhat overlapping techniques for probing deeply into mountains of e-learner data. This informal use of the terms glosses over the extensive data structures and innovate techniques used to do the probing. The common use of the terms generally refers to computational techniques applied in order to uncover patterns in huge data sets about online teaching and learning. The underlying techniques draw on a variety of sophisticated and ever-improving machine learning algorithms. Encompassing a wide range of goals and approaches, learning analytics and data mining of user activity in e-learning systems have become research fields in their own right in recent years (Siemens & Baker, 2012). Typical approaches focus on how to find patterns in learner online behavior. Arranging various patterns into groupings (e.g., based on the activities, roles, and timing involved) can shed light on issues such as how to evaluate student progress or recommend learning pathway options. The variety of

learning analytics and educational data mining investigations is also broad and ever increasing, but some of the better-known approaches include clustering, association analysis, and predictive analytics (Romero & Ventura, 2010).

As related fields, learning analytics and educational data mining also represent burgeoning research and policy areas where the teacher training and teacher research fields' traditional thought leadership and policy expertise will be much needed. A fair generalization can be made that much of the inquiry and practical wisdom developed so far center on applying computational techniques to large data sets about students. Applying tracking and data mining techniques in online teaching and learning contexts, learning analytics and educational data mining encompass a unique range of research questions and policy issues. Learning analytics and educational data mining efforts in secondary and tertiary education settings have served as the basis for discovering categories and characteristics in student enrollment patterns. In the context online learning environments, data mining projects have examined similarities across thousands of online sessions to reveal useful characteristic aspects of students' interaction with e-learning content as well (McGrath, 2009). The influence of learning analytics and educational data mining on secondary and tertiary education is potentially enormous. The easy response to this new technology, i.e., unwavering acceptance of it as a black box technology would be a tragic mistake in the face of required demand. With or without the teacher training program's involvement, many learning analytics and educational data mining-based attempts at creating metric-driven smart school will spring up in the coming decade. Within the context of online learning, an important set of strategy and policy considerations arises. With teaching and learning activities increasingly moving online, important research and policy questions surface as to how users are to be studied, how their usage patterns should be captured, how that user data will get analyzed, by whom and for what purposes.

3 Implications for Teacher Training Validity and Inferencing

With the early generation of learning analytics and educational data mining systems, important warnings have already been raised about both the myriad privacy concerns and the tremendous sociopolitical implications of the data mining revolution on a global scale. Comprehensive surveys of the privacy issues can be found in Ferguson (Ferguson et al., 2016). An overview of the critical data studies field is provided by Kitchin and Lauriault (2014) and Illiadis and Russo (2016). For education, some of the particularly salient concerns raised here include the ownership and commodification of learner data (Pardo & Siemens, 2014), governance and policy (Slade & Prinsloo, 2013), and the emerging data "divide" that mirrors the socioeconomic digital divide of previous decades (Dalton, Taylor, & Thatcher, 2016).

Meanwhile, even as we are rightly concerned about these critically important issues (e.g., confidentiality of learners' activities and the longer term data inequities), it is important in the near term to recognize as well a fundamental set of methodological problems within the emerging data sciences disciplines driving this movement. Namely, there is a significant methodological gap between the promise of the new technology and its ability to deliver reliable results. Learning analytics, educational data mining, and data science, in general, are beginning to experience growing pains as technology implementations move from the research environment to the real world. As recently acknowledged in a watershed report from the National Academies, the immaturity of data mining and data analytics as disciplines is a potential crisis if not quickly addressed. The data sciences, according to this report, are years away from being reliably principled reliability from an engineering perspective and conclusion validity from a statistical perspective (Jordan, 2013).

As a result, one immediate area in which learning analytics and educational research would benefit from more engagement from the fields of teacher training and teacher research would be in bringing statistical rigor to the information frameworks being deployed. Indeed, the common technical challenges that are bedeviling early generation learning analytics and educational data mining systems are age-old familiar issues for educational research and statistical inferencing: measurement error, sample size, over-fitting, etc. (Baker & Inventado, 2014). While e-commerce and social media system for search engines and recommender services may be able to tolerate high order error rates in their results, a system focusing on the fate and trajectory of individual student learners can scarcely tolerate fractional error rates. This typical challenge faced by designers of learning analytics and educational data mining systems stems in part from the relentless combining of disparate data sources—a technique that undergirds all web analytics technology. Digital systems cut across a wide range of teaching and learning activities in secondary and tertiary education today. The scope and reach of digital systems now increasingly extend to activities as they occur both inside and outside of physical classrooms, labs, and informal study areas. Electronic books, learning management systems, interactive student response systems, lecture capture systems, and digitally controlled smart classrooms are just a few examples of technology trends that potentially bring along with them an unprecedented amount of instrumentation quietly collecting lots of data about teacher and learner activities in and across these various spaces. In snapshots, these usage streams offer data that can be helpful for understanding and supporting teaching and learning. If combined across time and location, the varied data sources open windows onto even more interesting activity patterns and relations.

These mosaics, however, are very difficult to create and analyze in ways that meet traditional approaches to reliability and validity assumptions about data (Birgersson, Hansson, & Franke, 2016; Zhu et al., 2014; Doan, Domingos, & Halev, 2001). The reliability of traditional parametric statistical methods, for instance, requires as a starting point some assumptions about estimators and requirements about the probability distribution of the overall population from which data samples are drawn. In contrast, data mining approaches typically make no assumptions about models in the underlying data. Not making assumptions about models and distributions is

partly seen as a way of allowing for serendipity. The exploratory knowledge discovery nature of data mining is valued for finding hidden patterns. More practically, the application of traditional parametric methods to big data can make exploration infeasible, resulting in either the discarding of much data or a computational complexity that makes timely results prohibitive. So data mining approaches relax the rigorous requirements of traditional parametric methods as a necessary cost in reliability and controlling uncertainty of achieving good enough results in a timely fashion (Larose, 2007). As a consequence divining rods, many implementers stray from inferential rigor and resort instead to heuristic techniques. These heuristic techniques such as nearest neighbor machine learning algorithms for classifying data by membership into groups. As the algorithm “learns” from training set data, it improves in its ability to assign class membership at some practical level of reliability that is often quite functional and suitable for some applications, such as profiling users of an e-commerce system or selecting customers as the audience for a marketing campaign.

Where the risks and consequences involved in misclassifying some of the data are acceptable, data mining’s departure from traditional guidelines of reliability, error, and bias are deemed acceptable in some contexts (Glymour et al., 1997; Dasu & Johnson, 2003). Misclassifying a consumer for inclusion in a marketing campaign involves little impact. Someone getting a pop-up advertisement that turns out to be of no interest to them can dismiss it and move on. In contrast, misclassifying a learner regarding their progress in school may have a lasting impact. A student getting classified as needing remediation may find it very difficult to shake such a label (Prinsloo & Slade, 2016). While teaching itself involves plenty of informed guesses within the moment, the field of education has long embraced inferential methods for the many situations where informed guessing is not good enough. It is important, for example, to quantify certainty in deciding whether a learning outcome has been met, a new instructional method is effective, or a student should matriculate. The main and simplest point here is that basic notions of confidence intervals, sampling, and proportion estimates are already part of the traditional teacher training and teacher research toolboxes. The field of education can bring to the educational data mining and learning analytics conversations a balanced perspective on requirements for quantifying the degree of uncertainty and the use of statistical decision-making. As learning analytics and educational data mining are increasingly becoming available as mainstream research topics in educational research, there are plenty of opportunities to expand the focus to consider to engineer them better as reliable decision support systems (Pardo, 2014). Meanwhile, teachers, teacher training candidates, and teacher researchers alike must specifically develop critical and reflective perspective and stances toward these new technologies.

4 Implications for Teacher Research—More Theory, Thicker Description

As we've seen, teacher training programs and the field of teacher research need to become more critically engaged with learning analytics and educational data mining particularly regarding the reliability and validity of the answers being given. The second reason for critical engagement, we will see next, stems from the kinds of questions being asked. Many of the early generation systems developed and studied so far focused heavily on technology development and proof-of-concept prototypes, with the teaching and learning settings serving as mere background. Indeed, the educational questions, subjects, and issues in many studies, it seems, are chosen simply to provide algorithmic testbeds based on the convenient access to log data. As we will see, even in the case of production systems that have seen some success, the learning analytics and educational data mining approaches employed have demonstrated useful albeit very *narrow* insights: most commonly in detecting students who are in need of intervention or remediation.

This narrowness starts to make sense if we consider that typically is analyzed in early generation learning analytics and educational data mining systems: the so-called click streams left behind by students visiting, browsing, and interacting with e-learning content and tools. While the strength of the new technologies can be found in their ability to deal with huge and diverse data sets, a potential weakness stems from this same reliance on gathering pre-existing usage data. Behind the typical early generation learning analytics and educational data mining systems are evolving efforts to bring together more usage data regarding both source and volume. Most of these efforts, however, face practical hurdles: pulling together whatever usage data is available from disparate online tools and services and combining them by using loose-coupling and lightweight data standards. To accomplish these tasks, the functionality for combining and analyzing learning activity and learner information often gets boiled down to even simpler common denominators. Obviously, the scope of the patterns, arrangements, or groupings to be discovered depend heavily on the breadth and depth of the user activity streams in the original clickstream data.

Looking at some prominent research studies in the field, we can start to see the constraining effect of the data source availability. In the case of the Purdue University's Course Signals, for instance, the key data element used as a proxy for "effort" was simply the student's overall usage pattern in the course site within learning management system (Arnold and Pistilli, 2012). These traces of usage activity, combined with other educational analytics (e.g., test scores, GPA, standardized test scores, unit load, age, etc.) were mined to produce "actionable intelligence." Visualized in a rudimentary green, yellow, red dashboard rating for each student's potential risk of failure, the actionable information thereby gives instructors and support personnel high-level signals about student progress. The same constraining effect of the available data sources can also be seen in the units of analysis studied so far in the promising Open High School of Utah (OHSU) project, where learning analytics play a crucial role in mediating teacher and student interaction (Tonks, Weston, Wiley,

& Barbour, 2013). Given that students and teachers are not copresent in a physical school building, online analytics become essential in this virtual school situation for recording and monitoring individual student access to course materials, discussion forum activity, and their assessment results. In an online school, the volume from the various forms of user activity data captured grows by quickly by the day. In the context of the Moodle learning management system deployed for OHSU, instructors are provided with some monitoring capabilities as well as some predictive learning analytics about the students as derived from the thousands of hours of students accessing the virtual school's course sites and tools. Nevertheless, the breadth and quality of the data analytics here still depend on the what's available in the data source—in this case, Moodle activity logs.

A large cross-institutional open source project such as Moodle, for instance, involves scores of developers around the world over the years contributing to a shared code base. To facilitate distributed development, the design of the Moodle framework places minimal requirements on those who might want to create or integrate a new tool. By minimizing the overhead of tool creation and rewrites, however, the Moodle framework offers very little out-of-the-box functionality in the area of usage reporting, as Romero points out in his data mining study of Moodle use at the University of Cordoba (Romero, Ventura, & Garcia, 2008). The behind-the-scenes view of Moodle in operation reveals a piecemeal and heterogeneous affair. In particular, since responsibility for logging information about users' interaction within a running instance of Moodle is largely left up to individual tool developers, the usage data is inconsistent. In the case of OHSU, these limits have meant that learning analytics system is necessary but not sufficient for instructors in supporting teacher decision-making (Borup, Graham, & Drysdale, 2014). In the OHSU example, the deployment of these new technologies in a virtual school setting was shown to provide some benefits in narrow cases: monitoring, identifying students at risk, remediation, just-in-time alert systems, etc.

Of course, teaching and learning involve far more than just monitoring student presence and mitigating situations in which some students risk failing (Macfadyen & Dawson, 2010). First-generation learning analytics and educational data mining system have been shown to succeed in small online focused areas of early alert and remediation, but how to extend these new approaches to broader theoretical models and concerns of teaching and learning? Here, most observers do not yet have answers. For George Siemens, a major leader in the field, such issues are the main challenge for learning analytics and educational data mining if they are to survive. The next generation of learning analytics and educational data mining must focus aspects of pedagogy, he argues. To overcome the early generation limitation, argues Siemens, a new design approach for developing learning analytics and educational data mining for development must include learning from the start:

Some analytics techniques, such as early warning systems [12, 13], attention metadata [14], recommender systems [15], tutoring and learner models [16], and network analysis [17], are already in use in education. A few papers in LAK11 presented analytics

approaches that emphasized newer techniques, such as participatory learning and reputation mechanisms [18], recommender systems improvement [19], and cultural considerations in analytics [20]. Beyond these, however, there are limited first-generation LA techniques. The lack of defined identity of LA tools and techniques with an explicit learning focus is reflected in how analytics are described in papers and conference venues: “It’s like Shazam”, or “It’s like Amazon or Netflix”, or “It’s like Facebook friend recommendations”. This is not to criticize appropriating techniques from other fields for use in learning. Instead, it is a reflection that LA-specific approaches are still emerging and more research is required.

(Siemens, 2012, p. 6)

Siemens does not say how to achieve a more theory-driven approach. However, he does correctly pinpoint a key relation where many of these factors would come into play at the earliest stage of learning analytics and educational data mining systems design: the tensions between bottom-up approaches based on available data and top-down approaches based on theoretical inquiry. Some new set of design processes is needed, Siemens asserts, for balancing local needs against the top-down constraints. What Siemens has put his finger on here, a process by which system functionality and data source descriptiveness would be better shaped by theory-driven questions, rather than the reverse.

Wider recognition of the need for more theory-driven approaches has begun to emerge as the single most important concern of the new these new fields (Dawson, Mirriahi, & Gasevic, 2015). Regarding learning analytics and educational data mining study results connecting to theoretical models of teaching and learning, the constraints of the data sources have limited the scope and power even in the few studies that have attempted modest theoretical claims (Tempelaar, Rienties, & Giesbers, 2015; Pardos, 2015). So another reason for teacher training programs and the field of teacher research to become more directly engaged in the future development of learning analytics and educational data mining include the need for more theory-driven approaches in these new fields (Dawson et al., 2015).

5 Conclusion

This chapter has considered the future of the learning analytics and educational data mining. The two fundamental shortcomings of these new fields are the limited instructional models considered and the relative immaturity of these new technologies when viewed from traditional perspectives of inferencing. In terms of models of instruction, the barriers preventing these systems from developing deeper insights into the teaching and learning activities seem mundane but vexing: the limited data sources upon

which these systems can draw. In terms of the immaturity of these new technologies when viewed from traditional perspectives of inferencing and decision-making support, the potential for bias and inaccuracy in the algorithms involved is not merely an engineering problem. It points ahead to a perpetual need for transparency and openness so that algorithms are not concealed in proprietary black boxes where they might avoid scrutiny. Issues around the validity of inferential approaches employed and the narrowness of underlying data being mined point to political and policy questions that must be raised as learning analytics and data mining as decision-support systems are proposed for use in secondary and tertiary education. As we have seen, the challenges and issues seen in the early generation of learning analytics cannot simply be dismissed as growing pains.

This chapter has also pointed to the need for educational professionals to consider not only how such systems are designed and implemented, but also how they could be built better in the future. The influence of learning analytics and educational data mining on secondary and tertiary education is growing quickly. For the field of teaching, a passive response to this new technology, i.e., acceptance of it as a black box technology that cannot be questioned would be a mistake. Indeed, teacher training programs have before them a historic opportunity to influence fundamentally how learning analytics and educational data mining will be deployed and used. This is a role for which the teacher training and teacher research programs are uniquely suited: to influence research agendas, to form, fund, and nurture critical perspectives (Baeppler & Murdoch, 2010). Bringing to bear wisdom from a century's worth of scholarship on teaching and learning would well befit educational research and teacher training programs, given their long-standing leadership in researching and assessing technology initiatives in teaching and learning. An important technology convergence is at hand again, one that holds out the promise of tracking, monitoring, measuring, and adapting teaching and learning activity in schools as a means of designing and assessing instructions with adaptive personalization. The field of education could bring to the educational data mining and learning analytics development not only domain expertise but also a balanced perspective on grounding the risks around of statistical decision-making. Finally this central role of the education domain experts, in turn, would necessarily require that educators become more literate in data science as well.

With or without the engagement from the fields of teacher research and teacher training, many learning analytics and educational data mining-based attempts at creating metric-driven smart schools will spring up in the coming decade try to address secondary and tertiary schooling from the perspective of measurement, accountability, and access (Daniel, 2012). The teacher training and teacher research fields' traditional roles as thought leaders in educational research have stemmed historically from their methodological expertise in collecting, managing, and analyzing data about teaching and learning. Teacher training and teacher research fields should extend that tradition by contributing to the evaluation and design of these new systems, bringing along core expertise in methods of educational research and inferencing. Teacher training and teacher research fields also possess unique capacity as a leading contributor to educational policy. By engaging more directly with learning analytics

and educational data mining, the fields could develop teachers' critical literacy and expertise, while also shaping and advancing policies geared toward ensuring openness and transparency in how these new knowledge domains of learning analytics and educational data mining are implemented and managed. Professional educators in general also have a responsibility to serve policy advocates around best practices and watchdogs on the lookout for privacy and bias problems. These need already exist. Many more issues and opportunities will become known in the context of virtual school implementations. If professional educators take the leadership role in helping design and create model implementations of learning analytics and educational data mining, the fields of teacher training and teacher research would be in a strong position to ensure that the technology development and implementation are guided systematically by open debate, ethical policies, and grounded understanding of best practices.

References

- Arnold, K. E., Pistilli, M. D. (2012). Course signals at Purdue: Using learning analytics to increase student success. In S. Buckingham Shum, D. Gašević, R. Ferguson (Eds.), *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)* (pp. 267–270). New York: ACM.
- Bach, C. (2010). LA: Targeting instruction, curricula, and student support. In *Proceedings EISTA 2010 June 29—July 2, 2010*. Orlando, FL: International Institute of Informatics and Systemics.
- Baepler, P., & Murdoch, C. J. (2010). Academic analytics and data mining in tertiary education. *International Journal for the Scholarship of Teaching and Learning*, 4(2), 1–9.
- Baker, R. S., & Inventado, P. S. (2014). Educational data mining and learning analytics. In R. S. Baker & P. S. Inventado (Eds.), *Learning analytics* (pp. 61–75). New York, NY: Springer.
- Birgersson, M., Hansson, G., & Franke, U. (2016). Data integration using machine learning. In *2016 IEEE 20th International Enterprise Distributed Object Computing Workshop (EDOCW)*, Vienna, Austria.
- Blikstein, P. (2013). Multimodal learning analytics. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge*. New York: ACM.
- Borup, J., Graham, C. R., & Drysdale, J. S. (2014). The nature of teacher engagement at an online high school. *British Journal of Educational Technology*, 45(5), 793–806.
- Bramble, W., & Panda, S. (2008). Organizational and cost structures for distance and online learning. In: W. Bramble, & S. Panda (Eds.), *Economics of distance and online learning*. London & New York: Routledge.
- Brown, J., & Kurzweil, M. (2017). *The complex universe of alternative post-secondary credentials and pathways*. Cambridge, MA.: American Academy of Arts & Sciences.
- Carlsen, W. S. (2001). Domains of teacher knowledge. In J. Gess-Newsome & N. G. Lederman (Eds.), *Examining pedagogical content knowledge* (pp. 133–144). Dordrecht, Netherlands: Kluwer Academic.
- Cuban, L. (2001). *Oversold and underused: Computers in the classroom*. Cambridge, MA: Harvard University Press.
- Dalton, C. M., Taylor, L., & Thatcher, J. (2016). Critical data studies: A dialog on data and space. *Big Data & Society*, 3(1).
- Daniel, J. S. (2010). *Mega-schools, technology, and teachers: Achieving education for all*. London & New York: Routledge.

- Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox, and possibility. *Journal of Interactive Media in Education*.
- Daniel, B. (2015). Big data and analytics in tertiary education: Opportunities and challenges. *British Journal of Educational Technology*, 46(5), 904–920.
- Dasu, T., & Johnson, T. (2003). *Exploratory data mining and data cleaning*. New York: Wiley.
- Dawson, S., Mirriahi, N., & Gasevic, D. (2015). Importance of theory in learning analytics in formal and workplace settings. *Journal of Learning Analytics*, 2(2), 1–4.
- Doan, A., Domingos, P., & Halevy, A. Y. (2001). Reconciling schemas of disparate data sources: A machine-learning approach. *ACM Sigmod Record*, 30(2), 509–520.
- Ferguson, R., Hoel, T., Scheffel, M., & Drachler, H. (2016). Guest editorial: Ethics and privacy in learning analytics. *Journal of Learning Analytics*, 3(1), 5–15.
- Glymour, C., Madigan, D., Pregibon, D., & Smyth, P. (1997). Statistical themes and lessons for data mining. *Data Mining and Knowledge Discovery*, 1(1), 11–28.
- Iliadis, A., & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2).
- Jagdish, H. V. (2015). Big data and science: Myths and reality. *Big Data Research*, 2(2), 49–52.
- Jordan, M. I., et al. (2013). *Frontiers in massive data analysis*. Washington, D.C.: The National Academies Press.
- Jung, I. (2005). ICT-Pedagogy integration in teacher training: Application cases worldwide. *Educational Technology & Society*, 8(2), 94–101.
- Kenny, C. (2006). *Overselling the web: Development and the internet*. Boulder, CO: Lynne Rienner.
- Kleickmann, T., Richter, D., Kunter, M., Elsner, J., Besser, M., Krauss, S., et al. (2013). Teachers' content knowledge and pedagogical content knowledge the role of structural differences in teacher education. *Journal of Teacher Education*, 64(1), 90–106.
- Kozma, R. B. (2008). Comparative analysis of policies for ICT in education. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 1083–1096). New York, NY: Springer.
- Kitchin, R., & Lauriault, T. P. (2014). Towards critical data studies: Charting and unpacking data assemblages and their work. In J. Eckert, A. Shears, & J. Thatcher (Eds.), *Geoweb and big data* (The programmable city working paper 2; pre-print version of chapter to be published). University of Nebraska Press. Available at SSRN: <https://ssrn.com/abstract=2474112> (Forthcoming).
- Kumar, V. S., Somasundaram, T. S., Boulanger, D., Seanosky, J., & Vilela, M. F. (2015). Big data LA: A new perspective. In Kinshuk & Huang (Eds.), *Ubiquitous learning environments and technologies*. Berlin: Springer.
- Larose, D. T. (2007). *Data mining methods and models*. New York: Wiley.
- Long, P., & Siemens, G. (2011). Penetrating the fog: Analytics in learning and Education. *Educause Review*, 48(5), 31–40.
- 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.
- McGrath, O. (2009). Mining user activity data in tertiary education open systems: Trends, challenges, and possibilities. In T. Kidd (Ed.), *Handbook of research on technology project management, planning, and operations*. Hershey, PA: Information Science Reference.
- Moe, T., & Chubb, J. (2009). *Liberating learning: Technology, politics, and the future of American education*. San Francisco: Jossey-Bass.
- O’Neil, C. (2016). *Weapons of math destruction. How big data increases inequality and threatens democracy*. New York: Crown.
- Pardo, A. (2014). Designing learning analytics experiences. In *Learning analytics* (pp. 15–38). New York: Springer.
- Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450.
- Pardos, Z. A. (2015). Commentary on “Beyond time-on-task: the relationship between spaced study and certification in MOOCs”. *Journal of Learning Analytics and Knowledge*, 2(2), 70–74.
- Prinsloo, P., & Slade, S. (2016). Student vulnerability, agency, and learning analytics: An exploration. *Journal of Learning Analytics*, 3(1), 159–182.

- Romero, C., Ventura, S., & García, E. (2008). Data mining in course management systems: Moodle case study and tutorial. *Computers & Education*, *51*(1), 368–384.
- Romero, C., & Ventura, S. (2010). EDM: A review of the state-of-the-art. *IEEE Transaction on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, *40*(6), 601–618.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, *15*(2), 4–14.
- Siemens, G. (2012, April). Learning analytics: Envisioning a research discipline and a domain of practice. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge* (pp. 4–8). New York: ACM.
- Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In *Proceedings of the 2nd International Conference on learning analytics and knowledge* (pp. 252–254). New York: ACM.
- Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, *57*(10), 1510–1529.
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning Analytics in a data-rich context. *Computers in Human Behavior*, *47*, 157–167.
- Tonks, D., Weston, S., Wiley, D., & Barbour, M. K. (2013). “Opening” a new kind of school: The story of the Open High School of Utah. *The International Review of Research in Open and Distributed Learning*, *14*(1), 255–271.
- Zhu, H., Madnick, S. E., Lee, Y. W., & Wang, R. Y. (2014). *Data and information quality research: Its evolution and future*.