

What Does a Learning Analytics Practitioner Need to Know?

Leah P. Macfadyen
Faculty of Arts, The University of British Columbia,
1866 Main Mall, Vancouver, BC, V6T 1Z4, Canada
leah.macfadyen@ubc.ca

Abstract

The question captured in the title of this paper was asked by an audience member in a panel session at the 2014 *Learning Analytics Summer Institute*, hosted by the Society for Learning Analytics Research¹. More accurately, the question was: “What does a data scientist working in a learning context need to know?” One respondent quipped in response: “learn Python and learn R!”. Subsequent debate about this technically-oriented answer spilled over into an email thread in the *Learning Analytics Google Group*, a “news and discussion group for conceptual and practical use of analytics in education, workplace learning, and informal learning”, which currently has more than 1300 members worldwide. This paper briefly reviews the historical development of the field of learning analytics, in an effort to explain the perceived priority of technology skills. Selected contributions to the online discussion are offered, in the hope that they can inform ongoing efforts to develop a coherent, comprehensive and relevant learning analytics curriculum.

1. Introduction

Lang et al. [1] catalyzed discussion of the need for ‘a learning analytics curriculum’ by noting that the field of learning analytics is now being advanced not only through research and publication, but “in the classroom” - shorthand for the recent launch of an array of online and face to face courses and programs in learning analytics. Additional learning opportunities also include the now regularly occurring *Learning Analytics Summer Institute*² sponsored by the Society for Learning Analytics Research, as well as other ‘professionally minded projects’ such as the European *Learning Analytics Community Exchange* project (LACE)³.

A key goal of the LAK 2017 conference workshop “Building the Learning Analytics Curriculum” was to “seed a community that can communicate about problems of practice around the teaching of Learning Analytics” (p. 521). One fundamental challenge to the development of a coherent and mutually agreed upon learning analytics curriculum, however, may be disagreement within the ‘learning analytics community’ itself regarding “what a learning analytics practitioner needs to know”. It is common for learning analytics learning materials to be dominated by training in analytic techniques and tools, and for knowledge of learning theories, social theories and learning design to be absent or considered only peripherally.

This brief paper explores the origins and nature of this tension, and points towards a framework that may assist in development of a sufficiently interdisciplinary and diversified learning analytics curriculum.

¹ <http://solaresearch.org/>

² <https://solaresearch.org/events/lasi/>

³ <http://www.laceproject.eu/lace/>

2. Is Data Science the New Black?

Exploratory data analysis is part of the logic of analytics. I would argue that you can't do sophisticated exploratory data analysis without knowledge of python or r.

(Alfred Essa, Vice President, R&D and Analytics, McGraw-Hill Education)

2.1. Data Science Steals the Limelight: The Rise of Big Data

Ferguson [2] and Elias [3] have usefully charted the emergence of the field of learning analytics in the past fifteen years, giving us clues about how and why ‘technical expertise’ occupies a priority position in the minds of some practitioners. Learning analytics research has roots in a variety of fields, including business intelligence, web analytics, educational data mining and recommender systems – all of which critically depend on application of data skills. And while learning analytics might perhaps be characterized as a relatively small and specialized branch of the analytics tree, business intelligence - and ‘analytics’ more generally – have garnered a tidal wave of attention this decade, which shows no sign of abating.

In 2011, The McKinsey Global Institute published an influential report [4] which characterized effective analytic use of big data as the “next frontier for innovation, competition and productivity”, on a global scale. Their extensive review argued that effective use of big data had already and demonstrably allowed sectors such as marketing, sports, retail, health and technology to enhance productivity, systems and outcomes, and identified education as a ‘sector’ that lagged behind others in embracing the potential of analytics. The report also presented worrying projections that by 2018, the US alone might be facing a 50-60% gap in ‘deep analytical talent’ (read: technical talent) compared to projected demand. Such speculation (and such need) has driven the current massive proliferation of courses and programs in data science, business analytics, analytics, statistics, programming...and R and Python. For example, a 2016 ranking of ‘top 50 MOOCs by learner ratings’ [5] revealed that 40% are ‘technology’ courses – almost exclusively comprising courses in data analysis and programming topics. Seventeen of the top 50 MOOCs ranked by *enrollments* belong to this same group [6]. Universities are following close behind with new offerings of residential and online graduate degrees in data science and related topics.

Arguably, data science and analytics currently have the limelight, driving career advising, development of professional development opportunities, and professional education. As a consequence, a wave of new graduates and professionals with ‘data skills’ is entering the labour market, and with it the common assumption that such skills may simply be applied in any field.

2.2. Mining Educational Data

Did analysis of educational data really begin with the new wave of data scientists? To be fair, early educational analytics efforts predate the current (and largely business-driven) big data boom. Romero and Ventura [7] trace the origins of ‘educational data mining’ (EDM) back to 1995, when universities were starting to make use of learning technologies that generated increasingly detailed sets of log data of student-computer interaction. EDM is a subset of the larger field of data mining, “a field of computing that applies a variety of techniques (for example, decision tree construction, rule induction, artificial neural networks, instance-based learning, Bayesian learning, logic programming and statistical algorithms) to databases in order to discover and display previously unknown, and potentially useful, data patterns” [2]. This existence of this earlier computer-science-led approach to investigating educational data has doubtless also shaped perceptions that technical skills have priority.

In the past decade, EDM and learning analytics have rubbed shoulders, sometimes jostling uncomfortably for pre-eminence. Commentators e. g. [8] agree that these fields “share many attributes and have similar goals” – in particular, the goal of improving education. But efforts to differentiate the two often characterize EDM and LA as having different technological, ideological and methodological orientations. In particular, EDM approaches typically place much greater emphasis on “automated discovery” – that is to say, more reductive and technology-oriented modelling approaches that predict learner behaviour or outcomes, and guide automated adaptation of learning materials, without human involvement. Learning analytics approaches, on the other hand, are characterized as having a more holistic and systems-oriented approach. Learning analytics models are typically integrated into tools and reports whose goals are to inform and empower decision-makers in the learning context (instructors and learners). Even as learning analytics practitioners sometimes seek to distance themselves from ‘pure EDM’, however, techniques borrowed from EDM remain key tools in the learning analytics toolkit.

2.3. Data Skills: Necessary but Not Sufficient

...anyone can learn the tools, and many of the analytic skills transfer easily, but without subject matter expertise the results will be nonsensical and inapplicable in the real world...Good analytics grow out of subject expertise first and foremost. Handing me a hammer, even the best on the market, won't make me a good carpenter. Handing a programmer r won't make them a good data scientist.

Rebecca T. Barber, Statistician and Data Scientist, Arizona State University

What makes education perhaps a bit special is its complexity and how deceiving it looks from the outside. In my experience...when computer scientists approach education, for some obscure reason, they tend to see that the success is around the corner by using their expertise and finding the right tool combo and some basic knowledge of the context...We CS people are tinkers and when tinkering in any context, we need to have the right tools. We are obsessed with this (as it should be). I find value in knowing what tools are considered important to put in my backpack...But I need...the constant reminder that I still don't fully understand what happens in a learning context.

Abelardo Pardo, Lecturer, School of Electrical and Information Engineering
The University of Sydney, Australia

Critical to the evolution of learning analytics as a new field, however, has been the vital integration of social and pedagogical insights and theories into the endeavour. Signs of this ‘social and pedagogical turn’ can be detected in work dating from 2003 [2], as researchers began to situate their work within a constructivist pedagogical paradigm – a theoretical framework which holds that knowledge is constructed through social negotiation [9, 10]. Learning analytics studies have increasingly made use of methodologies that go beyond data mining and automated discovery, introducing approaches such as social network analysis, and, later, discourse analysis, natural language processing, multimodal learning analytics, and others. Importantly, learning analytics work has increasingly drawn on social and pedagogic theories, as the field has more explicitly articulated its focus as being understanding and optimizing *learning*.

It had become increasingly clear that while analytics skills are a necessary component of the work of the field, learning cannot simply be understood by algorithms alone. Organizers of the first *Learning Analytics and Knowledge* conference in Banff in 2011 emphasized the urgent need for a more nuanced, theory-driven and interdisciplinary approach to understanding learning, arguing that “technical, pedagogical, and social domains must be brought into dialogue”⁴.

Learning analytics is now a highly interdisciplinary field that draws on diverse literature from education, technology and the social sciences, with valuable contributions from researchers and practitioners who approach ‘learning’ from multiple perspectives (Table 1).

Table 1. Fields Contributing to Learning Analytics Research and Practice

<i>Technical/Analytic</i>	<i>Social Sciences/Education</i>
<ul style="list-style-type: none">• statistics• data visualization and visual analytics• educational data mining• computer science• machine learning• natural language processing• human-computer interaction• and others....	<ul style="list-style-type: none">• social sciences• education• (educational) psychology• psychometrics• cognitive science• educational technology• learning design• art and design• and others...

⁴ LAK’11. 1st International Conference on Learning Analytics and Knowledge 2011. <https://tekri.athabascau.ca/analytics/>

3. Connected specialization

Edu data science" is not a fruit borne from a single tree. Rather it is the result of grafting together multiple fields since currently there is no single educational or experiential pathway to cultivate the educational data scientist.

Phil Arcuria, Director of Research,
Glendale Community College, Arizona, USA

I'm becoming more convinced that the focus needs to be on connected specialization. We need complimentary, not duplicated, skills...the system works because we connected specialized skills sets rather than generalized skill sets. I imagine that LA implementations should fall on the side of specializations and team models.

George Siemens, Executive Director,
Learning Innovation and Networked Knowledge Research Lab,
University of Texas at Arlington, USA

We need to have a team of people with specialized skills (i.e., ML and DM algorithms and R/python tools, educational theory and instructional design knowledge, information visualization specialist) but we also need integrators, administrators and entrepreneurs in order to make LA endeavors successful.

Vitomir Kovanovic,
Doctoral Student, University of Edinburgh, Scotland

Where does this leave us? If we acknowledge the highly interdisciplinary nature of the field, how can we identify core credentials, competencies and curriculum needs? Phil Arcuria (see quote, above) usefully proposed a framework for training learning analytics specialists. All should, he argues, “have a basic working knowledge and/or skill level (broadly defining both terms without taxonomic distinction) in all the areas and a high level of expertise in at least one area”. The foundational elements he suggests – encompassing research design and methods, domain knowledge, and foundations of learning theories – are included in Table 2.

Additional knowledge and skillsets valuable to learning analytics teams are suggested by common themes in the current literature. James Williamson (Office of Information Technology, University of California, Los Angeles, USA) and others noted, for example, the importance of “knowledge/awareness of privacy and other data ethics issues”.

Moreover, in recent years, the challenges of integrating learning analytics at the institutional level, and of effecting system-wide change in complex systems such as higher education, have also garnered attention [11-14]. “Seven things not to do” when seeking to transform patterns of student success at your institution, include overlooking the reality that a campus is an ecosystem and seeking to “fix only one thing”, importing external solutions under the assumption that “one-size-fits-all”, forgetting that campus culture shapes outcomes dramatically, overlooking implementation issues, and insisting on top-down implementation [15]. Further elaboration of barriers to, and successful models for, institutional transformation suggest that effective learning analytics teams also need members with a deep understanding of institutional systems and complexity, and who have expertise and skills in managing organizational change. As doctoral student Vitomir Kovanovic noted, above, effective integration of learning analytics into different educational context calls not only for specialized analysts and learning theorists, but for individuals with an understanding of the learning analytics field, who are also effective managers – characterized in Adizes’ model of prototypical management styles as “producers”, “integrators”, “administrators” and “entrepreneurs” [16].

Happily, development of learning materials that span many of these topics is already underway. The *Handbook of Learning Analytics* [17], newly launched in 2017, offers a comprehensive survey of relevant learning analytics topics contributed by researchers and practitioners in the field. What remains, perhaps, is the juggling act of achieving balance, recognizing that it is neither realistic nor necessary to expect all learning analytics practitioners to have an expert understanding and skill level in all areas.

One final suggestion may keep us honest:

... if I were to reframe the question as "what one skill/experience will help to make a good learning analytics person better", I'd answer "teach a class"...Nothing helps one understand the interactions in a class better (especially in an online class) than teaching. You start to understand patterns, student motivations, terminology, and lots of little nuances that all help to fill in bits of color in the picture.

Mike Sharkey,
President & Founder, Blue Canary Data & Analytics

Table 2. Proposed Areas of Necessary Knowledge for Learning Analytics Practitioners

-
- **Learning theories:** Knowledge of the general theories of learning.
 - **Subject matter expertise:** Knowledge of the relevant educational context, and its data.
 - **Computer science/Programming:** Knowledge of foundational concepts, an understanding of the basics underlying all/most programming languages, e.g. operands, operators, Boolean logic, understanding of the logic of core routines, e.g. loops, IF statements, etc.).
 - **Statistics/Analytics:** Knowledge of traditional (e.g., NHST, frequentist, etc.) and non-traditional (machine learning, Bayesian, etc.) techniques.
 - **Research design:** Knowledge of different types of research designs, and advantages/disadvantages of each.
 - **Communication skills:** Ability to effectively communicate complex material in a digestible form via various media.
 - **Information Technology:** Knowledge of databases, how data are stored and retrieved, basic querying fundamentals, e.g. joins, keys, etc.
 - **Data ethics and privacy issues:** Principles of data privacy and data governance, local and national laws and policies, the risks and benefits of learning analytics tools and strategies.
 - **Fundamentals of organizational change:** An understanding of systems models and approaches to organizational change.
-

4. Acknowledgments

My thanks to LASI 2014 participants and Learning Analytics Google Group members who have contributed to this discussion and furthered thinking about the development of a relevant learning analytics curriculum.

5. References

- [1] C. Lang, S., Teasley, J. Stamper. Workshop: Building the Learning Analytics Curriculum. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference, Vancouver, British Columbia, Canada, March 13 - 17, 2017 (pp. 520-521). ACM New York, NY, USA, 2017
- [2] R. Ferguson. Learning analytics: Drivers, developments and challenges. *International Journal of Technology Enhanced Learning*, 4(5/6): 304-317, 2012.
- [3] T. Elias. Learning analytics: The definitions, the processes, and the potential. <https://landing.athabascau.ca/file/download/43713>, 2011.
- [4] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, A. H. Byers. Big data: The next frontier for innovation, competition and productivity. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation, 2011
- [5] D. Shah. Class Central's Top 50 MOOCs of All Time. <https://www.class-central.com/report/top-moocs/>, July 19, 2016
- [6] Online Course Report. The 50 Most Popular MOOCs of All Time. <http://www.onlinecourserreport.com/the-50-most-popular-moocs-of-all-time/>, retrieved June 30, 2017.
- [7] C. Romero, S. Ventura. Educational data mining: a survey from 1995 to 2005. *Expert Systems with Applications*, 33(1):135-146, 2007.

- [8] G. Siemens, G., R. S. J. D. Baker. Learning analytics and educational data mining: Towards communication and collaboration. In S. Buckingham Shum, D. Gasevic, R. Ferguson, Proceedings, Learning analytics and knowledge. Banff, AB, Canada: ACM, 2011.
- [9] R. Aviv, Z. Erlich, G. Ravid, A. Geva. Network analysis of knowledge construction in asynchronous learning networks. *Journal of Asynchronous Learning Networks*, 7(3):1–23, 2003.
- [10] M. De Laat, V. Lally, L. Lipponen, L., R.-J. Simons. Analysing student engagement with learning and tutoring activities in networked learning communities: a multi-method approach. *International Journal of Web Based Communities*, 2(4): 394–412, 2006.
- [11] L. P. Macfadyen, S. Dawson, S. Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3):149–163, 2012.
- [12] R. Ferguson, L. P. Macfadyen, D. Clow, B. Tynan, S. Alexander, S. Dawson. Setting learning analytics in context: Overcoming the barriers to large-scale adoption. *Journal of Learning Analytics*, 1(3):120-144, <http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/4077/4421>, 2014.
- [13] L. P. Macfadyen, S. Dawson, A. Pardo, D. Gašević. Embracing Big Data in Complex Educational Systems: The Learning Analytics Imperative and the Policy Challenge. *Research & Practice in Assessment*, 9:17-28, <http://www.rpajournal.com/dev/wp-content/uploads/2014/10/A2.pdf>, 2014.
- [14] L. P. Macfadyen. Overcoming Barriers to Educational Analytics: How Systems Thinking and Pragmatism Can Help. *Educational Technology*, 57 (1):31-39, 2017.
- [15] G. L. Mehaffy. Student Success: Seven Things Not to Do. In K. Kruger, R. Martin, G. L. Mehaffy, J. O'Brien, Student Success: Mission-Critical. *EDUCAUSE Review*, 52(3), <http://er.educause.edu/articles/2017/5/student-success-mission-critical>, May/June 2017.
- [16] I. Adizes. Management/Mismanagement Styles: How to Identify a Style and What To Do About It. Santa Barbara, California: Adizes Institute Publishing, 2004.
- [17] C. Lang, G. Siemens, A. Wise, D. Gašević (Eds.). The Handbook of Learning Analytics (First Edition). Society for Learning Analytics Research, <https://solaresearch.org/hla-17/>, 2017.