Learning Analytics: Envisioning a Research Discipline and a Domain of Practice

George Siemens
Technology Enhanced Knowledge Research Institute
Athabasca University
Edmonton, AB T5J 3S8
780-421-5841

gsiemens@athabascau.ca

ABSTRACT
Learning analytics are rapidly being implemented in different educational settings, often without the guidance of a research base. Vendors incorporate analytics practices, models, and algorithms from datamining, business intelligence, and the emerging "big data" fields. Researchers, in contrast, have built up a substantial base of techniques for analyzing discourse, social networks, sentiments, predictive models, and in semantic content (i.e., "intelligent" curriculum). In spite of the currently limited knowledge exchange and dialogue between researchers, vendors, and practitioners, existing learning analytics implementations indicate significant potential for generating novel insight into learning and vital educational practices. This paper presents an integrated and holistic vision for advancing learning analytics as a research discipline and a domain of practices. Potential areas of collaboration and overlap are presented with the intent of increasing the impact of analytics on teaching, learning, and the education system.

Categories and Subject Descriptors
J.1 [**Administrative Data Processing**] *Education*; K.3.1 [**Computer Uses in Education**] *Collaborative learning, Computer-assisted instruction (CAI), Computer-managed instruction (CMI), Distance learning*

General Terms
Algorithms, human factors

Key Words
Learning Analytics, Theory, Research, Practice, Collaboration, Ethics, Data Integration

1. INTRODUCTION
Learning analytics (LA) is a young and developing concept. Reflection is warranted on how to position early developments for long-term viability and positive impact of LA on learning and teaching. Of critical importance is increased dialogue between researchers and practitioners in order to guide the development of new tools and techniques for analytics.

LAK12: *2nd International Conference on Learning Analytics & Knowledge*, 29 April – 2 May 2012, Vancouver, BC, Canada

It is uncertain at this stage whether LA will develop as a distinct field of study or whether analytics techniques will be subsumed into existing research fields. Regardless of the long-term trajectory of LA, a research base is already rapidly developing. The Learning Analytics and Knowledge conference registrations doubled from 2011 to 2012 (from 99 to over 200) and submissions for review increased from 38 to 90. Numerous special issues of academic journals (see http://www.ifets.info/ and http://sloanconsortium.org/publications/jaln_main) indicate that LA is gaining interest in different research fields. Additional indicators of LA's continued growth can be found in government reports [1] and numerous EDUCAUSE papers [2].

The first international LA conference in Banff in 2011, Learning Analytics and Knowledge (LAK), emphasized the importance of bridging computer sciences and social sciences:

> Advances in knowledge modeling and representation, the semantic web, data mining, analytics, and open data form a foundation for new models of knowledge development and analysis. The technical complexity of this nascent field is paralleled by a transition within the full spectrum of learning (education, work place learning, informal learning) to social, networked learning. These technical, pedagogical, and social domains must be brought into dialogue with each other to ensure that interventions and organizational systems serve the needs of all stakeholders. [3]

To date, this social and technical connection has been largely positive, but needs continued focus to advance LA's impact on learning. With the significant increase in interest in data and analytics, as indicated by conferences, journals, grant funding opportunities, and growing vendor base, educators and researchers have an opportunity to influence the development of analytics in education.

LA is a sprawling term, at times referring to complex predictive models and at other times to routine tasks such as classroom allocation and energy conservation. The Society for Learning Analytics (SoLAR, http://solaresearch.org/) emphasizes the learner in its definition: "Learning analytics is 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" [4].

This definition includes techniques such as predictive modeling, building learner profiles, personalized and adaptive learning, optimizing learner success, early interventions, social network analysis, concept analysis, and sentiment analysis. For a more detailed overview of the development and approaches of LA, Ferguson has a comprehensive account [5].

The International Educational Datamining Society (IEDMS, http://www.educationaldatamining.org) community is progressing on a parallel path to SoLAR in developing techniques and approaches to understand the learning process through analytics and data [6]. IEDMS (EDM at the time) initiated a workshop series in 2005 and in 2008 hosted its first international conference. Together, SoLAR and IEDMS seek to discover new insights into learning and new tools and techniques so that those insights impact the activity or practitioners in primary, secondary, and higher education, as well as corporate learning.

2. The Gap

The research and practice gap, prominent in numerous fields including corporate finance [7] and social sciences [8], is evident within LA. The work of researchers often sits in isolation from that of vendors and of end users or practitioners. This gap is challenging as it reflects a broken cycle of communication and interaction between empirical research and how those findings are translated into practice.

Researchers, especially in technology fields, are not found only in university laboratories. Many software companies invest heavily in research, creating and ensuring protection of their intellectual property (IP). Researchers in the LA domain explore learning through the lens of data and analytics and share findings with peers through publications and conferences so that other researchers can build on discoveries. Sharing and disseminating findings, algorithms, and new tools through scholarly discourse is vital for innovation.

The practitioners who use the tools and techniques of researchers and vendors are educators, designers, data scientists, trainers, managers, and university administrators. The translation of research into practice occurs through the knowledge development and teaching roles of universities and through the risk-taking activities of corporations as they develop products and services. The university sector is a vital contributor to the knowledge economy, providing states, provinces, and nations with competitive advantages [9]. Researchers' activities contribute to the development of tools and techniques that influence corporate activity. Vendors serve as a bridge between researchers and practitioners as they translate research findings into software or product offerings. Data are constantly generated. Collection and analysis in LA do not have a natural beginning or conclusion. Each analysis activity potentially feeds into practice, and each practitioner act can serve as a new data point. Early indications from vendors who are developing analytics software suggest that findings will be treated as proprietary and will not be made available to other researchers. The growing prominence of protected IP can hinder iterative and rapid improvements to LA techniques.

3. Why is this an Important Conversation?

LA researchers risk losing relevance in the rapidly developing world of corporate learning analytics companies (i.e., corporate networks that integrate content companies with companies that offer adaptation and personalization platforms). Researchers should recognize that much of the innovation in LA happening in the vendor space. Unfortunately, many vendor-driven innovations are closed and do not meet the basic needs of researchers: open, testable, accessible, and improvable algorithms and tools.

A few tools, such as R (http://www.r-project.org/), have been developed with openness in mind (even though they were not developed as LA tools). Researchers have already acquiesced significant ground in building an open analytics platform to the vendor community. A healthy vendor community is vital for LA to make an impact, but researchers need access to their tools and data in order to validate and test findings. How are various factors weighted in an algorithm? Are the concepts being analyzed the right ones? Can researchers adjust the algorithms of vendor tools to conduct experiments of other factors that might impact learning?

Researchers require transparency and the ability to expose their work to scrutiny of the broader community. Practitioners need tools that are easy to use and that provide a positive end-user experience. This will often require that much of the technical functionality of analytics be hidden as an end-user layer guides practitioners in how to interact with the data. Vendors develop and commercialize tools and services, informed by research, that allow for broad deployment. Practitioners, in their use of tools and techniques, can inform both researchers and vendors.

Researchers seek to understand LA from multiple perspectives: learners, institutions, and effectiveness. Corporations do not share this obligation. Context influences the nature and type of analytics. Understanding and disseminating analytics practices and algorithms will assist researchers in building better models. When these models are open, customization based on context is possible. Increased dialogue between researchers, vendors, and practitioners will unlikely solve the research-practice gap, but it will raise awareness of the needs of each entity and generate a sense of the important role that each plays.

4. Holistic and Integrated Research/ Practice Relationships

The U.S. Department of Education has stated that the "next 5 years will bring an increase in models for collaboration between learning system designers, researchers, and educators" [1]. Such models would include participation from numerous stakeholders in the analytics process: learners, faculty, departments, institution, researchers, corporations, and government funders. A holistic view of LA includes a broad spectrum of educational activity, including the full student experience: pre-enrolment in university, learning design, teaching/learning, assessment, and evaluation.

It is possibly futile to layout the direction needed in an emerging field. The momentum in LA is significant and rapid changes are difficult to track and it is even more difficult to trace their trajectory. Yet in spite of the uncertainty around analytics in education, a few considerations are important for researchers, practitioners, and vendors in order to position LA for long-term success. To advance as a field, LA researchers and practitioners need to address the following: (a) development of new tools, techniques, and people; (b) data: openness, ethics, and scope; (c) target of analytics activity; and (d) connections to related fields and practitioners.

4.1. Development of tools, techniques, and people

Three areas of development are needed to drive the adoption of analytics in education: new tools and techniques, the practitioner experience, and the development of analytics researchers.

Analytics tools and techniques that focus on the social pedagogical aspect of learning are required. Numerous techniques have been developed outside of the education system, often from business intelligence research. In other instances, the tools used for analysis have not scaled with the increase in data size or sophistication of analytics models. For example, discourse analysis has a long history [10] in educational research. However, dramatic increases in the size of discourse data sets, such as those generated in large online courses, can overwhelm manual coding. In response, automated analysis of discourse [11] builds on existing models while scaling to accommodate the analysis of larger data sets.

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.

The second aspect of development, the practitioner experience, focuses on the end user experiences of analytics tools and techniques. Researchers and vendors present practitioners with tools that are too complex. First-generation LA tools involve researchers actively engaged with practitioners, providing oversight, guidance, and support. As LA begin to inform more of the education system, such as curriculum design, advising learners, and pedagogical practice, practitioners using the tools will have varying technical skillsets. Intuitive and easy-to-use tools are important in involving greater numbers of educators. The next generation of tools must be designed to serve a dual purpose: context-sensitive help and guidance for non-technical users and an accessible technical layer that allows more advanced users to interact directly with data and to tweak and adjust analysis models.

The final area of development centers on the capacity of practitioners and researchers. Practitioner skills and knowledge are being developed through traditional educational programs. A few universities, such as North Caroline State University (http://analytics.ncsu.edu/) and Northwestern (http://www.analytics.northwestern.edu/), have started offering masters programs in analytics. These programs are focused on business intelligence, but many techniques are transferrable to education. The capacity for analytics deployment requires the development analysts and practitioners at masters and doctoral levels. Certificate programs for university leaders and administrators are not yet available. Professional development programs are anticipated to address this need in the near future.

For researchers, capacity for LA research is being created in various fields: computer science, statistics, programming, network analysis, and psychology of education. In order to bring these fields in dialogue with each other, research labs will need to be developed. A distributed online research lab has been proposed to help develop analytics students and researchers: http://www.solaresearch.org/lab/. If the current trajectory of LA development continues, it is reasonable to expect traditional research labs to emerge that serve a similar role of bringing specialized analytics fields in relation to each other.

4.2. Data: Openness, ethics, scope

As analytics derives from data, it is not surprising that many outstanding concerns in LA centre on data. Foundational issues of data quality, ethics of use, scope of analytics activity, data standards, and integration data sets will continue to occupy a large part of the conversation. Additionally, big data [21] raises the prospect of new research methods [22]. It is conceivable that future research models, especially in complex domains like education, will be based on analytics rather than existing research models. Conceivably, an explosion in learning sciences research will result.

Ethics, learner rights, and data ownership are prominent topics. Early attempts at clarifying data ownership recognize the need for learner control [23]. Mismanagement of the messaging around ethics in analytics can result in learner, and even broader public, pushback to LA as a field. Analytics researchers, practitioners, vendors, and educational systems have a responsibility communicate clearly and transparently the scope and role of an LA deployment. A proactive stance of transparency and recognition of potential learner and educator unease of analytics may be helpful in preventing backlash.

Human-contributed feedback or corrective options are also required. When systems develop profiles and models of learners, anomalies and errors can be expected. Recommender systems, for example, may provide personalized content to a learner based on learners who share similar profiles. This content may not be helpful to a specific learner. End users, when given options of correcting or "teaching" recommender systems, can improve personalization.

The data silos that exist in universities, schools, and organizations present a profound challenge for both researchers and practitioners. Data integration from multiple sources can improve the accuracy of a learner profile and subsequent adaptation and personalization of content. Sharing data across silos addresses a weakness in existing LA activity: data is too centric to learning management systems (LMS) and student information systems (SIS). Learner activity captured in these two systems comprises only a fraction of the learning process. The inclusion of data from other sources, such as mobiles, sensors, physical world data, advising, and the use of university resources such as libraries and tutors, will result in a more complete learner profile. New data sets create exponentially to building learner profiles: LMS data, combined with SIS data and the social media profile of a learner, affords analytics opportunities that far exceed single data points.

Ideally, an integrated analytics system would allow data and analytics layering: using multiple data sets and analytics techniques in a single interface for visualizing and presenting data to practitioners [24]. When these analytics tools are learner facing, learners can gain insight into their habits and the impact of their learning activities, thereby improving their self-awareness and self-regulation.

Two final considerations regarding data include: semantic data and real-time analytics. Extending analytics to include the role linked data and semantic technologies will enable better relationships between social and computing systems [25]. Semantic content (i.e., intelligent curriculum) will enable computers to provide personalized content to learners. Learner activity and their evolving profile can be constantly matched with the knowledge architecture of a particular domain and learning resources provided to fill any knowledge gaps.

Secondly, once analytics are conducted in (near) real-time, learners will receive notification of conceptual errors earlier than they currently do when the educators marks exams or essays. For educators, real-time analytics and visualization will identify challenges facing different learners based on concept comprehension (as a result of lectures, labs, or simulations) or through sentiment analysis (i.e., self-confidence) of discourse.

4.3. Target of analytics activity

Analytics are frequently cast as primarily technical or statistical activities. A transition in analytics from a technical orientation to one that emphasizes sensemaking, decision-making, and action [26] is required to increase interest among educators and administrators. This transition in no way minimizes the technical roots of analytics; instead, it recasts the discussion to target problems that are relevant to practitioners instead of researchers.

A second needed transition is one that moves LA research and implementation from at-risk identification to an emphasis on learner success and optimization [27, 28]. Identifying at-risk learners has been, and will continue to be, an important deliverable for LA. College dropouts are a concern facing universities and society (and obviously the learners). However, identifying at-risk learners is a small aspect of what analytics can do to improve education. Through social network analysis and content recommender systems, automated marking, improved learner self-awareness, and real-time feedback for educators, the learning process can be significantly optimized.

4.4. Connections to related fields and practitioners

Improving connections with related fields of research, such as machine learning, educational data mining, learning sciences, psychology of learning, and statistics, is vital. The pieces that define analytics are scattered across these fields. Working with and sharing distributed knowledge is challenging but important [29].

In addition to connecting various research domains, LA must consider how it interacts with education systems, leaders and other stakeholders. It is necessary to promote realism around what learning analytics are and what they are able to accomplish. Resolving, or at least suitably responding to, concerns about the inability of data to capture complex social processes such as learning are also required. In his presentation in SoLAR's open online course, Campbell [30] emphasizes the limitations of analytics to measure complex processes such as learning. Nuanced and thoughtful messaging should address both the hype and buzz around analytics as well as voices that discount LA are nothing new.

Research organizations and industry associations are the likely agents to serve this society-facing role. For example, both SoLAR and IEDMS are well positioned in this regard. Association publications are still needed that target administrators and policy makers. These reports could include annual state of the industry analysis to communicate how the LA ecosystem is evolving in terms of analytics adoption, implementation models, and the vendor community.

5. Conclusion

Theoretically, LA has potential to dramatically impact the existing models of education and to generate new insights into what works and what does not work in teaching and learning. The results are potentially transformative to all levels of today's education system. For example, as models of personalization and adaptation of learning develop, do we still need a course model in higher education? Are schools and universities allocated resources in those areas that make the biggest impact? How will we learn in a networked, distributed, mobile, and analytics-driven system?

Answering these questions through research, and then translating those findings into practice, requires an evaluation of the current state of LA and the challenges that need to be addressed. These challenges currently involve the development of new tools, techniques, and people; resolving data concerns such openness, ethics, and the scope of data being captured; enlarging and transitioning the target of analytics activity; and improving connections to related fields. The task is significant and difficult, but well worth embracing given the large potential benefit of an integrated and holistic LA researcher-practitioner model.

6. REFERENCES

[1] M. Bienkowski, M. Feng, and B. Means. *Enhancing teaching and learning through educational datamining and learning analytics* (Draft): page 45, 2012. Available from <http://evidenceframework.org/wp-content/uploads/2012/04/EDM-LA-Brief-Draft_4_10_12c.pdf> [accessed April 15, 2012]

[2] *EDUCAUSE: Learning analytics: 69 resources*, 2012. Available from <http://www.educause.edu/Resources/Browse/LearningAnalytics/39193> [accessed April 15, 2012]

[3] G. Siemens. *About: Learning Analytics & Knowledge: February 27-March 1, 2011 in Banff, Alberta:* para. 4. July 22, 2010. Available from <https://tekri.athabascau.ca/analytics/> [accessed April 15, 2012]

[4] Society for Learning Analytics Research. *About,* 2012. Available from http://www.solaresearch.org/about/ [accessed April 15, 2012]

[5] R. Ferguson. *The state of learning analytics in 2012: A review and future challenges*, 2012. Available from <http://kmi.open.ac.uk/publications/pdf/kmi-12-01.pdf> [accessed April 15, 2012]

[6] R. S. J. Baker and K. Yacef. *The state of educational data mining in 2009: A review and future visions,* 2009. Available from <http://www.educationaldatamining.org/JEDM/images/articles/vol1/issue1/JEDMVol1Issue1_BakerYacef.pdf> [accessed April 15, 2012]

[7] E. A. Trahan and L. J. Gitman. Bridging the theory-practice gap in corporate finance: A survey of Chief financial officers. *Quarterly review of economics and Finance*, 35: 73-87, January 1995.

[8] P. Lather. Research as praxis. *Harvard Educational Review*, 56(3): 257-278, Fall 1986.

[9] D. F. Shaffer and D. J. Wright. *Higher education. A new paradigm for economic development: How higher education institutions are working to revitalize their regional and state economies*, March 2010. Available from <http://www.rockinst.org/pdf/education/2010-03-18-A_New_Paradigm.pdf> [accessed April 15, 2012]

[10] G. Brown and G. Yule. *Discourse analysis.* London, England; Cambridge University Press, 1983. Available from <http://www.cambridge.org/gb/knowledge/isbn/item1130157/?site_locale=en_GB> [accessed April 15, 2012]

[11] A. De Liddo, S. Buckingham Shum, S., I. Quinto, M. Bachler, and L. Cannavacciuolo. *Discourse-centric learning analytics*. <http://dl.acm.org/citation.cfm?id=2090120&CFID=82269174&CFTOKEN=35344405> [accessed April 1, 2012]

[12] J. P. Campbell, W. B. Collins, C. Finnegan, and K. Gage. *Academic analytics: Using the CMS as an early warning system*. Paper presented at WebCT Impact 2006. Chicago, IL, July 2006.

[13] L. P. Macfadyen and S. Dawson. Mining LMS data to develop an "early warning system" for educators: A proof of concept. *Computers & Education, 54*(2): 588-599, February 2010.

[14] E. Duvall. Attention please!: learning analytics for visualization and recommendation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*: pages 9-17, March 2011. doi:10.1145/2090116.2090118

[15] Y. H. Cho, J. K. Kim and S. H. Kim. A personalized recommender system based on web usage mining and decision tree induction. *Expert Systems with Applications*, 23(3): 329-42, September 2002.

[16] P. Brusilovsky. Adaptive hypermedia: From intelligent tutoring systems to web-based education. *User Modeling and User-Adapted Interaction*, 11(1-2):87-110, 2001.

[17] M. Newman. *Networks: An introduction*. New York, NY: Oxford University Press, 2010.

[18] D. Clow and E. Makriyanis (2011) iSpot analysed: Participatory learning and reputation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*: pages 34-43, March 2011. doi:10.1145/2090116.2090121

[19] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari and E. Duval. Dataset-driven research for improving recommender systems for learning. In *Proceedings of the 1st International Conference Learning Analytics & Knowledge*: pages 44-53, March 2011. Available from <http://dl.acm.org/citation.cfm?id=2090122&CFID=77412043&CFTOKEN=45426814> [accessed April 3, 2012]

[20] R. Vatrapu. Cultural considerations in learning analytics. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*: pages 127-133, March 2011. doi:10.1145/2090116.2090136

[21] F. X. Diebold. *"Big data" dynamic factor models for macroeconimic measurement and forecasting* [online], 2000. Available from <http://www.ssc.upenn.edu/~fdiebold/papers/paper40/temp-wc.PDF> [accessed June 4, 2011]

[22] C. Anderson. *The end of theory: The data deluge makes the scientific method obsolete*, June 2008. Available from <http://www.wired.com/science/discoveries/magazine/16-07/pb_theory> [accessed April 15, 2012]

[23] Office of Science and Technology Policy, Executive Office of the President. *Unlocking the power of education data for all Americans* [Fact sheet], January 19, 2012. Available from <http://www.whitehouse.gov/sites/default/files/microsites/ostp/ed_data_commitments_1-19-12.pdf> [accessed April 4, 2012]

[24] D. Suthers and D. Rosen. A unified framework for multi-level analysis of distributed learning. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge:* pages 64-74, March 2011. doi:10.1145/2090116.2090124

[25] J. Hendler and T. Berners-Lee. *From the semantic web to social machines: A research challenge for AI on the world wide web*, November 2009. doi.org/10.1016/j.artint.2009.11.010 [accessed April 15, 2012]

[26] G. Siemens. *Sensemaking: Beyond analytics as a technical activity.* April 11, 2012. Available from <http://www.educause.edu/ELI124/Program/GS01> [accessed April 15, 2012]

[27] A. W. Astin. Degree attainment rates at American colleges and universities: Effects of race, gender, and institutional type. Los Angeles, CA: Higher Education Research Institute, University of California, 1996.

[28] V. Tinto. *Leaving college: Rethinking the causes and cures of student attrition*, 2nd edition. Chicago, IL: The University of Chicago Press, 1993.

[29] G. Stasser and W. Titus. Pooling of unshared information in group decision making: Biased information sampling during discussion. *Journal of Personality and Social Psychology*, 48(6): 1467-1478, June 1985. Available from http://www.citeulike.org/group/2546/article/1398512 [accessed April 15, 2012]

[30] G. Campbell. *Here I stand*. Presentation to Learning and Knowledge Analytics Open online course, 2012 (http://lak12.wikispaces.com/). Session recording available from <https://sas.elluminate.com/p.jnlp?psid=2012-03-01.1231.M.0728C08DFE8BF0EB7323E19A1BC114.vcr&sid=2008104>