Chapter 1: Theory and Learning Analytics

Simon Knight, Simon Buckingham Shum

Connected Intelligence Centre, University of Technology Sydney, Australia
DOI: 10.18608/hla17.001

ABSTRACT

The challenge of understanding how theory and analytics relate is to move “from clicks to constructs” in a principled way. Learning analytics are a specific incarnation of the bigger shift to an algorithmically pervaded society, and their wider impact on education needs careful consideration. In this chapter, we argue that by design – or else by accident – the use of a learning analytics tool is always aligned with assessment regimes, which are in turn grounded in epistemological assumptions and pedagogical practices. Fundamentally then, we argue that deploying a given learning analytics tool expresses a commitment to a particular educational worldview, designed to nurture particular kinds of learners. We outline some key provocations in the development of learning analytic techniques, key questions to draw out the purpose and assumptions built into learning analytics. We suggest that using “claims analysis” – analysis of the implicit or explicit stances taken in the design and deploying of technologies – is a productive human-centred method to address these key questions, and we offer some examples of the method applied to those provocations.

Keywords: Theory, assessment regime, claims analysis

In what has become a well-cited, popular article in *Wired* magazine, in the new era of petabyte-scale data and analytics, Anderson (2008) envisaged the death of theory, models, and the scientific method. No longer do we need to create theories about how the world works, because the data will tell us directly as we discern, in almost real time, the impacts of probes and changes we make.

This high profile article and somewhat extreme conclusion, along with others (see, for example, Mayer-Schönberger & Cukier, 2013), has, not surprisingly, attracted criticism (boyd & Crawford, 2011; Pietsch, 2014).

Educational *researchers* are one community interested in the application of “big data” approaches in the form of learning analytics. A critical question turns on exactly how theory could, or should shape research in this new paradigm. Equally, a critical view is needed on how the new tools of the trade enhance/constrain theorizing by virtue of what they draw attention to, and what they ignore or downplay. Returning to our opening provocation from Anderson, the opposite conclusion is drawn by Wise and Shaffer (2015, p. 6):

> What counts as a meaningful finding when the number of data points is so large that something will always be significant? [...] In sum, when working with big data, theory is actually more important, not less, in interpreting results and identifying meaningful, actionable results. For this reason we have offered Data Geology (Shaffer, 2011; Arastoopour et al., 2014) and Data Archeology (Wise, 2014) as more appropriate metaphors than Data Mining for thinking about how we sift through the new masses of data while attending to underlying conceptual relationships and the situational context.

Data-intensive methods are having, and will continue to have, a transformative impact on scientific inquiry (Hey, Tansley, & Tolle, 2009), with familiar “big science” examples including genetics, astronomy, and high energy physics. The *BRCA2 gene*, *Red Dwarf* stars, and the *Higgs bosun* do not hold strong views on being computationally modelled, or who does what with the results. However, when *people* become aware that their behaviour is under surveillance, with potentially important consequences, they may choose to adapt or distort their behaviour to camouflage activity, or to game the system. Learning analytics researchers aiming to study learning using such tools must do so aware that they have adopted a particular set of lenses on "learning" that amplify and distort in particular ways, and that may unintentionally change the system being tracked. Researchers should stay alert to the emerging critical discourse around big data in society, data-intensive science broadly, as well as within education where the debate is at a nascent stage.

Let us turn now to *educators and learners*. The potential of learning analytics is arguably far more significant than as an enabler of data-intensive educational research, exciting as this is. The new possibility is that educators and learners – the stakeholders who constitute the learning system studied for so long by researchers – are for the first time able to see their own processes and progress rendered in ways that until now were the preserve of researchers outside the system. Data gathering, analysis, interpretation, and even intervention (in the case of adaptive software) is no longer the preserve of the researcher, but shifts to embedded sociotechnical educational infrastructure. So, for *educators and learners*, the interest turns on the ability to gain insight in a timely manner that could improve outcomes.

Thus, with people in the analytic loop, the system becomes reflexive (people change in response to the act of observation, and explicit feedback loops), and we confront new ethical dilemmas (Pardo & Siemens, 2014; Prinsloo & Slade, 2015). The design challenge moves from that of modelling closed, deterministic systems, into the space of "wicked problems" (Rittel, 1984) and complex adaptive systems (Deakin Crick, 2016; Macfadyen, Dawson, Pardo, & Gašević, 2014). As we hope to clarify, for someone trying to get a robust measure of "learning" from data traces, such reflexivity will be either a curse or a blessing, depending on how important learner agency and creativity are deemed to be, how fixed the intended learning outcomes are, whether analytical feedback loops are designed as interventions to shape learner cognition/interaction, and so forth.

Our view is that it is indeed likely that education, as both a research field and as a professional practice, is on the threshold of a data-intensive revolution analogous to that experienced by other fields. As the site of political and commercial interests, education is driven by policy imperatives for "impact evidence," and software products shipping with analytics dashboards. While such drivers are typically viewed with suspicion by educational practitioners and researchers, the opportunity is to be welcomed if we can learn how to harness and drive the new horsepower offered by analytics engines, in order to accelerate innovation and improve evidence-based decision-making. Systemic educational shifts are of course tough to effect, but could it be that analytics tools offer new ways to evidence, at scale, the kinds of process-intensive learning that educators have long argued for, but have to date proven impractical? Exactly what one seeks to do with analytics is at the heart of this chapter.

To design analytics-based lenses – with our eyes wide open to the risks of distorting our definition of "learning" in our desire to track it computationally – we must unpick what is at stake when classification schemes, machine learning, recommendation algorithms, and visualizations mediate the relationships between educators, learners, policymakers, and researchers. The challenge of understanding how theory and analytics relate is to move "from clicks to constructs" in a principled way.

Learning analytics are a specific incarnation of the bigger shift to an algorithmically pervaded society. The frame we place around the relationship of theory to learning analytics must therefore be enlarged beyond considerations of what is normally considered "educational theory," to engage with the critical discourse around how sociotechnical infrastructures deliver computational intelligence in society.

The remainder of the chapter argues that by design – or else by accident – the use of a learning analytics tool is always aligned with *assessment regimes*, which are in turn grounded in *epistemological assumptions* and *pedagogical practices*. Moreover, as we shall explain, a long history of design thinking demonstrates that designed artifacts unavoidably embody implicit values and claims. Fundamentally then, we argue that deploying a given learning analytics tool expresses a commitment to a particular educational worldview, designed to nurture particular kinds of learners.

THEORY INTO PRACTICE

In an earlier paper (Knight, Buckingham Shum, & Littleton, 2014) we put forward a triadic depiction of the relationship between elements of theory and practice in the development of learning analytic techniques, as depicted in Figure 1.1 (we refer the reader to this paper for further discussion of the depicted relationships). Our intention was to illustrate the tensions and inter-relations among the more or less theoretically grounded stances we take through our pedagogic and assessment practices and policies, and their underlying epistemological implications and assumptions.

The use of a triangle highlights these tensions: that assessment *can* be the driving force in how we understand what "knowledge" is; that assumptions about pedagogy (for example, a kind of folk psychology; Olson & Bruner, 1996) influence who we assess and how; that assessment and pedagogy are sometimes in tension, where the desire for summative assessment overrides pedagogically motivated formative feedback; and that drawing alignment between one's epistemological view (of the nature of knowledge) and assessment or pedagogy practices is challenging – relationships between the three may be implied, but they are not *entailed* (Davis & Williams, 2002). Of course, other visualizations might be imagined, and the theoretical and practical purposes for which such heuristics are devised is important to consider. To give two examples, we have considered versions of the depiction in which: 1) assessment and pedagogy are built on the foundation of epistemology (in a hierarchical structure), and 2) are brought into alignment in a Venn diagram structure, with greater overlap implying a greater complementarity of the theorized position.

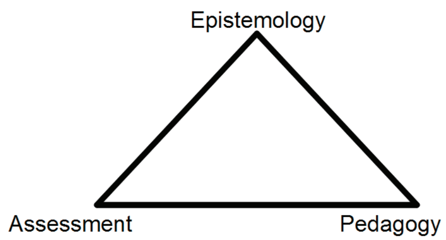

Figure 1.1. The Epistemology–Assessment–Pedagogy (EPA) triad (Knight et al., 2014, p. 25).

Learning analytics, as a new form of assessment instrument, have potential to support current educational practices, or to challenge them and reshape education; considering their theoretic-positioning is important in understanding the kind of education systems we aim for. For example, learning analytics could have the potential to 1) marginalize learners (and educators) through the transformation of education into a technocratic system; 2) limit what we talk about as "learning" to what we can create analytics for; and 3) exclude alternative ways of engaging in activities (that may be hard to track computationally), to the detriment of learners. Algorithms may both ignore, and mask some key elements of the learning process. The extent to which analytics can usefully support educators and learners is an important question. These are pressing issues given the rise of learning analytics, and increasing interest in mass online education at both the pre-university and university levels (e.g., the growing interest in MOOCs).

EPA PROVOCATIONS

Expanding on this prior work, the rest of the chapter aims to illustrate the application of our approach, with the aim of providing actionable guidance for those developing learning analytics approaches and tools. To do this, we have developed a set of provocations centred on the triad of epistemology, pedagogy, and assessment.

We use these provocations to illustrate how the implicit "claims" made by a learning analytics tool can be deconstructed. The approach invites reflection on the affordances of the tool's design at different levels (including data model, learner experience, and learning analytics visualization).

Computer-supported learning – individual or collaborative – covers a huge array of learning contexts. Such tools support many forms of rich learner interaction with peers and resources, which are implicit claims about learning. However, the emergence of computational analytics enables designers – and by extension the artifacts – to value certain behaviours above others, namely, those logged, analyzed, and rendered visible to some stakeholder group. The implicit claim is that these are particularly important behaviours. We measure what we value.

We provide a set of "six W" questions to be considered in the development of learning analytics. Of course, across these questions, there is overlap, and any one question might be expressed in multiple ways. The intention is neither to prescribe these as the only questions to be asked, nor that within each element of the triad *only* particular questions should apply. As the descriptions of the provocations make clear, within each facet of the triad, multiple theoretical questions can and should be asked. Rather, we hope to provide heuristic guidance to readers in developing their own analytics.

Epistemology – What Are We Measuring?

The first provocation invites the analytic designer to consider what "knowledge" looks like within the analytic approach being developed, asking, *What are we trying to measure*? We pose this question to prompt consideration of the connection between a conceptual account of the object of measurement (the knowledge being assessed), and a practical account of the methods and measures used to quantify activity and outputs within particular tasks. Asking *What are we trying to measure*? encourages us to consider our learning design, the skills and facts we want our students to learn, and what it means for students to "come to know." This is a question of epistemology; it concerns the nature of the constructs, why they "count" as knowledge, the evidentiary standard and kind required for a claim of knowledge to be made.

This knowledge might be of a more broadly propositional kind (sometimes characterized as "knowledge that," characterized as recall of facts), a more broad set of skills and characteristics (sometimes characterized as "knowledge how," for example, the ability to write an essay), or dispositions to act in particular ways (for example, as those dispositions recently discussed as epistemic virtues in epistemology). Evidentiary standards and types concern the warrants indicative of knowledge, for example, whether knowledge can be conceptualized in terms of unitary propositions that may be recalled more or less appropriately within particular contexts, whether knowledge of something entails the ability to deploy it in some context, the kinds of justification and warrant (and the skills underlying these) that cement some claim as knowledge, and so on. These are – implicitly or explicitly – the targets of our measurement.

Epistemology – How Are We Measuring?

Closely related to this conceptual question regarding the epistemological status of the object of analysis is a question regarding our access – as researchers and educators – to that knowledge. This is a question regarding the epistemological underpinning of our research and assessment methods. There is a rich literature on the various epistemological concerns around quantitative and qualitative research methods, with a growing specific interest in digital research methods. In addition, there is a focused literature in the philosophy of assessment, exploring the epistemological concerns in assessment methods (Davis, 1999). Across this literature, issues concerning the subjectivity of approaches, and the ability of methodologies to give insights, are central. The question invites considerations regarding the ways in which analytic methods imply particular epistemologies. Note that this is not just a question of the reliability of our assessment methods, but concerns the ability of approaches to speak to an externally knowable world (and the nature of that world).

Pedagogy – Why is this Knowledge Important to Us?

The development of analytic approaches in learning contexts involves making decisions about what knowledge will, and will not, be focused on; to measure what we value rather than value merely that which is easily measured (Wells & Claxton, 2002). This is, of course, in addition to a conceptual account of the nature of that knowledge. These decisions in part relate to debates around the kinds of important (or powerful) knowledge in society (see, for example, Young & Muller, 2015) and the role of knowledge-based curricula, including discussions around employability (or the balance of vocational and liberal educational aims), 21st-century skills, and so on. This question asks, *Why does this analytic matter to educators and learners*?

Answering this question might in part be salient to the kind of learning theory that the analytic sits within; to instrumental aims regarding the analytic's contribution to particular skills (perhaps employability skills); or university compliance (for example, reporting requirements). It might also relate to pedagogic aims such as the support of particular groups of students, and so on.

Pedagogy – Who is the Assessment/ Analytic For?

Extending the concern with the nature of the object of assessment above, is a further concern regarding the target of the analytic device, provoking the question, *Who is the analytic for*? In the development of analytic devices (and assessments more broadly), we should consider who the target of the device is, whether it supports teachers, parents, students, or administrators in understanding some aspect of learning. Is the analytic designed to provide insight at a macro (government, institutional), meso (school, class), or micro (individual student or activity) level (Buckingham Shum, 2012), and are there insights across these levels that can be effectively made sense of by all stakeholders (Knight, Buckingham Shum, & Littleton, 2013)?

This question regards the desire for analytic insights at multiple levels of a system, and the ability of individual analytic approaches (including their outputs in various forms, such as dashboards) to support the following: 1) individual students in developing their learning; 2) educators in developing their own practice and in targeting their support at individual student needs; and 3) administrators in understanding how cohorts are developing and their organizational needs. As Crick argues in this handbook, a complex systems conception of analytics for different levels in the learning system, spanning from private personal data through to shared organizational data, implies different rationalities and authorities to interpret at the different levels (Deakin Crick, 2017).

This question raises a parallel concern regarding the ethical implications in developing analytics that (explicitly or implicitly) target particular groups. This concern is at two levels. First, analytics that require particular forms of technology or participation may create new divides between student cohorts, or entrench existing divides. Second, there is an ethical concern regarding the use of student data by institutions, particularly where specific consent is not given, where no direct learning gain is directed to those students. This second issue is a particular consideration in cases wherein student data may be used largely to reduce institutional costs or the level of support given to particular students.

Assessment – Where Does the Assessment Happen?

Obvious though this is, we note that assessment always takes place in a physical location, in response to particular task demands, in a sociocultural context, with a particular set of tools. Contrast an individual pen-and-paper exam in a silent hall with 300 peers, with an emergency response simulation on the ward, with tackling a statistics problem in a MOOC, with conflict resolution in relationship counselling. For each context, we must ask not only if the assessment is meaningful, but also to what extent meaningful computational analytics can be designed to add value.

Moreover, we should also consider the ways in which the assessment biases particular kinds of response – in sometimes-unintended ways. For example, particular groups of students, or kinds of knowledge, might be privileged over others through the design of assessment contexts with a very narrow definition of achievement; through requirements for behaviours that not all students might engage in; through the use of technologies that unfairly assume socioeconomic means; or through separating assessment from the applied context in which an expertise can be displayed authentically.

Across assessments, we should also consider the ways in which the particular systems shape the data obtained – note this is a practical concern regarding the reliability and validity of methods, rather than the related epistemological concern raised above. For example, technologies are mediational tools, which shape the ways in which people interact with each other and the world around them, and hence, the activity they measure. This is true both of the specific technologies, and the task design used in assessments; for example, the use of "authentic" assessments provides a different range of possible responses than more traditional pen-and-paper assessments of various kinds.

Assessment – When Does the Assessment, and Feedback, Occur?

A final consideration relates to the temporal context of learning analytics, asking when the assessment and feedback cycle occurs. This provocation is intended to prompt consideration of the formative or/and summative nature of the learning analytic; whether or not a particular technology provides after-the-fact or real-time feedback, and whether this feedback is intended to provide a scaffold or model for current behaviour, is targeted at future behaviour and learning, or is just intended as a feedback mechanism on prior work (which may not be covered again).

In our earlier paper, we made a distinction between the metaphors of biofeedback and diagnostic learning analytics. The intention here was to draw a distinction between formative and summative assessments (respectively). However, while the analogy can be drawn, of course systems that provide real-time feedback can be summative in nature, and can take on a "monitoring" role towards some end-point. In addition, diagnosis does not have the finality perhaps implied by the analogy – diagnosis provides insight into what is going wrong, which may be actionable by the "patient" (student) or "doctor" (educator). Instead, then, the focus should be on whether the analytic device is targeted at a summative snapshot perspective on student learning and monitoring towards that end, or instead, targeted at development and improvement over time.

CONCLUSION

Through the provocations, we have drawn attention to the ways in which analytic approaches and artifacts subscribe to particular perspectives on learning: they implicitly *make claims* across the EPA triad, and the provocations drawn from them.

Tools can be used in many ways, and should not be isolated from their context of use. Interactional affordances, like beauty, are to some degree "in the eye of the beholder." We offer these provocations as a pragmatic tool for thinking, for designers, educators, researchers, and students – whether considering how one currently makes use of analytical tools, how one might do in the future, or indeed when designing new tools for new contexts. We propose that it is productive to consider these provocations in order to reflect on the EPA claims being made through the deployment of a learning analytic tool within a given context.

REFERENCES

Anderson, C. (2008, 23 June). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*. http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Arastoopour, G., Buckingham Shum, S., Collier, W., Kirschner, P., Knight, S. K., Shaffer, D. W., & Wise, A. F. (2014). Analytics for learning and becoming in practice. In J. L. Polman et al. (Eds.), *Learning and Becoming in Practice: Proceedings of the International Conference of the Learning Sciences* (ICLS '14), 23–27 June 2014, Boulder, CO, USA (Vol. 3, pp. 1680–1683). Boulder, CO: International Society of the Learning Sciences

boyd, danah, & Crawford, K. (2011, September). *Six provocations for big data*. Presented at A *Decade in Internet Time: Symposium on the Dynamics of the Internet and Society*. http://ssrn.com/abstract=1926431

Buckingham Shum, S. (2012). *Learning analytics policy brief*. UNESCO. http://iite.unesco.org/pics/publications/en/files/3214711.pdf

Davis, A. (1999). *The limits of educational assessment*. Wiley.

Davis, A., & Williams, K. (2002). Epistemology and curriculum. In N. Blake, P. Smeyers, & R. Smith (Eds.), *Blackwell guide to the philosophy of education*. Blackwell Reference Online.

Deakin Crick, R. (2017). Learning analytics: Layers, loops, and processes in a virtual learning infrastructure. In C. Lang & G. Siemens (Eds.), *Handbook on Learning Analytics* (this volume).

Hey, T., Tansley, S., & Tolle, K. (Eds.). (2009). *The fourth paradigm: Data-intensive scientific discovery*. Redmond, WA: Microsoft Research. http://research.microsoft.com/en-us/collaboration/fourthparadigm/

Knight, S., Buckingham Shum, S., & Littleton, K. (2013). Collaborative sensemaking in learning analytics. CSCW *and Education Workshop (2013): Viewing education as a site of work practice, co-located with the 16th ACM Conference on Computer Support Cooperative Work and Social Computing* (CSCW 2013), 23 February 2013, San Antonio, Texas, USA. http://oro.open.ac.uk/36582/

Knight, S., Buckingham Shum, S., & Littleton, K. (2014). Epistemology, assessment, pedagogy: Where learning meets analytics in the middle space. *Journal of Learning Analytics*, 1(2). http://epress.lib.uts.edu.au/journals/index.php/JLA/article/view/3538

Macfadyen, L. P., Dawson, S., Pardo, A., & Gašević, D. (2014). Embracing big data in complex educational systems: The learning analytics imperative and the policy challenge. *Research & Practice in Assessment*, 9, 17–28.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live work and think*. London: John Murray Publisher.

Olson, D. R., & Bruner, J. S. (1996). Folk psychology and folk pedagogy. In D. R. Olson & N. Torrance (Eds.), *The Handbook of Education and Human Development* (pp. 9–27). Hoboken, NJ: Wiley-Blackwell.

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. http://doi.org/10.1111/bjet.12152

Pietsch, W. (2014). Aspects of theory-ladenness in data-intensive science. Paper presented at the 24th Biennial Meeting of the Philosophy of Science Association (PSA 2014), 6–9 November 2014, Chicago, IL. http://philsci-archive.pitt.edu/10777/

Prinsloo, P., & Slade, S. (2015). Student privacy self-management: Implications for learning analytics. *Proceedings of the 5th International Conference on Learning Analytics and Knowledge* (LAK '15), 16–20 March, Poughkeepsie, NY, USA (pp. 83–92). New York: ACM.

Rittel, H. (1984). Second generation design methods. In N. Cross (Ed.), *Developments in design methodology* (pp. 317–327). Chichester, UK: John Wiley & Sons.

Shaffer, D. W. (2011, December). Epistemic network assessment. Presentation at the National Academy of Education Summit, Washington, DC.

Wells, G., & Claxton, G. (2002). *Learning for life in the 21st century: Sociocultural perspectives on the future of education*. Hoboken, NJ: Wiley-Blackwell.

Wise, A. F. (2014). Data archeology: A theory informed approach to analyzing data traces of social interaction in large scale learning environments. *Proceedings of the Workshop on Modeling Large Scale Social Interaction in Massively Open Online Courses at the 2014 Conference on Empirical Methods in Natural Language Processing* (EMNLP 2014), 25 October 2014, Doha, Qatar (pp. 1–2). Association for Computational Linguistics. https://www.aclweb.org/anthology/W/W14/W14-4100.pdf

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics*, 2(2), 5–13.

Young, M., & Muller, J. (2015). *Curriculum and the specialization of knowledge: Studies in the sociology of education*. London: Routledge.