



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Data power in education

Exploring critical awareness with the 'Learning Analytics Report Card'

Citation for published version:

Knox, J 2017, 'Data power in education: Exploring critical awareness with the 'Learning Analytics Report Card' ', *Television and New Media*, vol. 18, no. 8, pp. 734-752. <https://doi.org/10.1177/1527476417690029>

Digital Object Identifier (DOI):

[10.1177/1527476417690029](https://doi.org/10.1177/1527476417690029)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Television and New Media

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



Data Power in Education: exploring critical awareness with the 'Learning Analytics Report Card' (LARC)

Abstract

The burgeoning field of learning analytics (LA) is gaining significant traction in education, bolstered by the increasing amounts of student data generated through educational software. However, critical discussions of LA are in short supply. Drawing on work in the cultural studies of data and critical algorithm studies, this paper begins by examining three central issues: the distancing and 'black boxing' of LA disciplinary practices, the mythologizing of objective data, and the concern for future prediction. The second section describes the design and implementation of the 'Learning Analytics Report Card' (LARC), a pilot project that sought to develop experimental approaches to LA. As such, rather than seeking to simply produce analytics, the LARC attempted to foster critical awareness of computational data analysis amongst teachers and learners.

Introduction

As online distance learning, Massive Open Online Courses (MOOCs), and the provision of digital services more generally continue to thrive at higher education institutions, ever-increasing amounts of student data are being generated. Principally, in-house learning management systems (LMSs) capture vast amounts of information about the activities and behaviours of learners. This presents an exciting opportunity for universities: to not only work at the forefront of digital educational provision, but also to develop new insights, methods and pedagogical practices from the large quantities of data such projects generate. In the last three years, the New Media Consortium Horizon reports have consistently predicted significant impact from educational data mining and analytics (Johnson *et al.* 2014, 2015), with the latest version forecasting sector adoption for 'LA'

within the year (Johnson *et al.* 2016). Indeed, this term has emerged as a catch-all for any current data analysis approaches in education, although precise explanations of 'LA' are still being defined by the growing research community (Cooper 2012, Clow 2013). Global organisations are gaining traction, most notably the Society for Learning Analytics Research (<https://solaresearch.org/>), establishing increasingly high-profile research exchanges and international conferences. Initiatives supported by the Bill and Melinda Gates Foundation in the US are advancing broad approaches to computational data analysis¹, and large cross-Europe research projects are attracting significant funding².

The currency of LA is also demonstrated by the proliferation of courses purporting to teach the various techniques and skills foundational to the field. High-profile Massive Open Online Courses (MOOCs), include 'Practical Learning Analytics' from the University of Michigan (<https://www.coursera.org/course/pla>) on the Coursera platform, or 'Data, Analytics and Learning' from the University of Texas Arlington (<https://www.edx.org/course/data-analytics-learning-utarlingtonx-link5-10x>) delivered through edX. More formal higher educational offerings include an MSc in Learning Analytics from Teachers College Columbia (<http://www.tc.columbia.edu/human-development/learning-analytics/>), a Graduate Certificate from the liberal arts college Brandeis University (<http://www.brandeis.edu/gps/future-students/learn-about-our-programs/learning-analytics.html>), and 'Learning Analytics in the Knowledge Age' from the University of Minnesota (<http://meefen.github.io/blog/2015/01/27/learning-analytics-course/>). LA are also becoming establishing as integral parts of educational software, and recent years have seen software companies releasing analytics tools to accompany their LMS or MOOC platforms (Clow 2013).

However, this 'atmosphere of simultaneous hype and legitimate innovation' (Boellstorff and Maurer 2015, 2) presents considerable challenges for education. While 'big data' and its analysis continue to be publicised across academic, government and corporate research agendas, critical and

interdisciplinary approaches to educational data analysis are in short supply. The field of LA remains grounded in narrow disciplinary areas. Proponents often promise a fundamental disruption of education (Siemens 2013), yet draw exclusively from quantitative methodologies and theories from the learning sciences (Gasevic *et al.* 2015). This is coupled with complex technical methods and understandings from computer science disciplines that remain inaccessible to many in education.

The purpose of this paper is therefore to highlight key critical issues of power in computational analysis, and to bring different disciplinary perspectives – principally from the cultural studies of data (e.g. Andrejevic *et al.* 2015) and in particular critical algorithm studies in education (e.g. Williamson 2015) - to bear on the promise of LA. The cultural studies of data highlight the ways analytic technologies are ‘techno-economic constructs whose operations have important implications for the management of populations and the formation of subjects’ (Andrejevic *et al.* 2015, 380). This work foregrounds the ways in which culture is not only embedded in such systems, but is also reshaped through the power of algorithms (Beer 2009), which increasingly engage in ‘the sorting, classifying and hierarchizing of people, places, objects and ideas’ (Striphas 2015, 396). While critical algorithm studies have attested to this influence on social interaction (Gillespie 2014), and politics (Tufekci 2015), educational research is also addressing the penetration of algorithms into many aspects of teaching, learning and educational governance (see Perrotta and Williamson 2016).

Drawing on this work, the first section outlines three interrelated perspectives on LA. These highlight tendencies towards methodological estrangement and a ‘distancing’ from practices of teaching and learning; the assumption and privileging of objectivity; and the fixation on future prediction. The second section will describe the development and implementation of a small-scale LA project that sought to explore some of the critical questions around data analysis in education. Entitled the Learning Analytics Report Card (LARC), the project investigated ways of offering student

participation and choice in the analytics process, and the attempt to foster critical awareness of data capture and analysis amongst teachers and learners.

'Blackboxing' Analytics and educational (dis)empowerment

The information supplied by LA is often claimed to inform educational decision-making (Brown 2012, Siemens 2013, Norris and Baer 2013, McKenney and Mor 2015), and specifically to 'empower' educators in their practices (Camps-Febrer 2012, Clow 2013, ECAR-ANALYTICS Working Group 2015). However, the conditions for this 'empowerment' (attributed to teachers and students alike) are premised on the maturity of LA as a distinct discipline (Siemens 2013), and the development of specific expertise in areas outside of educational practice. A key point of definition in LA reveals this division. Whether termed 'actionable intelligence' (Campbell *et al.* 2007), 'data driven decision-making' (Barneveld *et al.* 2012), or 'fundamentally an interventionist science' (Rogers 2015), a clear boundary seems to be drawn between data analysis and educational response: the computational process remains hidden, while teachers and students are encouraged to react only to the results. While the need for interpretation of data is acknowledged here, those producing the analytics and those deciphering the results tend to operate in isolation. Clow defines this as a 'cycle', which 'starts with learners, who generate data, which is processed in to metrics, which are used to inform interventions, which in turn affect learners' (Clow 2013, 685). However, the point at which the processing of metrics takes place, the loop appears to divert into a different territory, one in which the underlying process is hidden from the view of teachers and students. It is the point of 'intervention' that is often emphasised, rather than the computational process that happens before it. Proposing 'actionable insights', Cooper suggests that 'analytics is concerned with the potential for practical action rather than either theoretical description or mere reporting' (2012, 4). Significantly then, the field of LA appears, not to deny the effects that data might have 'in the world', but rather

to actively encourage them, and to emphasise 'action' as the central concern. Cooper considers LA to be:

much more about a personal and organisational perspective on using data for decision-making and action-planning and less about how it is processed in a computer; evaluating, planning and doing are human activities. (Cooper 2012, 7)

The distinction being made here is therefore between active human intervention and passive non-human data processing. However, this reveals the problematic 'blackboxing' of LA – in other words the propensity to 'focus only on its inputs and outputs and not on its internal complexity' (Latour 1999, 304). For teachers, students and academic administrators, the actual processes of data analysis are largely hidden. Any educational 'action' or decision-making' is limited to responding to the results.

In practice, LA increase the disciplinary and spatial complexity of education: the processing of educational data tends to 'take place' amongst different academic communities, as well as private organisations (see Perrotta and Williamson 2016), and in computer laboratories, distributed server architectures and networks. However, this space of software, algorithms, and databases, alongside various human actors, is too often abstracted from the activities of education itself: student data is submitted, and teachers are encouraged to respond to the ensuing 'intelligence', without engaging in the underlying processes, and disciplinary cultures, with which it has been produced. As discussed in relation to the terms above, educational 'action' and 'decision-making' tend to be tolerated only *after* data has been collected, analysed and presented, limiting any sense of 'empowerment' to a reaction, rather than a position of knowing *how* new insights are generated. What is overlooked here is the substantial interest and investment that education might have in the *production* of LA. Rather than simply 'immaterial' - abstracted for the assumed advantage of educational actors - the

significant political and economic pressures that drive the data analysis agenda (for example, disproportionately towards assessment – see Clow 2013), as well as the numerous decisions and processes that determine the computational procedures themselves, may be of critical concern to educators and students alike. This is not to ignore the related tendencies for managerialism that increasingly pervade the neoliberal model of the university.

Furthermore, the institutional interest in these methods mean that ‘learning analytics is often applied outside an explicit research context; practitioners then have the responsibility to ensure that their practice meets those ethical standards’ (Clow 2013). boyd and Crawford have also highlighted the inability of ethics committees and institutional review boards to understand the nuanced issues of privacy brought about in big data analysis (2012). In other words, the ‘action’ and ‘decision-making’ previously described is often undertaken without adequate awareness of the ethical context of data processing. In the new ethical terrain of big data analysis, boyd and Crawford propose a notion of ‘accountability’, premised on the idea of being able to take responsibility for the way data is used (2012). However, where disciplinary divisions mask the processing of data, responsibility is problematically opaque, and to emphasise the educational ‘intervention’ merely leaves teachers, students, and institutional administrators accountable for the consequences.

For Clow, the solution is not for LA to become more transparent as a discipline, but rather for educators to become more educated:

The opportunity afforded by learning analytics is for educators to refuse to be overawed by the process, to understand the tools and techniques, their strengths and limitations, and to use that understanding to improve teaching and learning.

(Clow 2013, 693)

This seems to overlook the significant demands placed on teachers and students to understand the complex procedures - such as cluster analysis (see Perrotta & Williamson 2016) - that often comprise data analysis, pressures that will only increase as the technology develops. Andrejevic *et al.* also highlight the inaccessibility and lack of transparency inherent in the study of data (2016). Principally however, what these claims of empowerment appear to neglect is the ability to *not* use the results of data analytics in educational decision-making, and to resist being held accountable for such a choice. If LA were indeed capable of 'empowering' the teachers, administrators and students using it, one might assume their educational *autonomy* to be increased. However, it would be easy to assume that LA will lead to precisely the opposite: teachers and students being bound to act upon the results, and to be liable for not doing so. As Willis suggests, compelling educators to merely respond to the outputs of LA places the technological intervention firmly in the position of power and agency (2014).

The 'mythology' of objective data

The distinction between the (non-human) processing of data and the (human) educational intervention assumed in the literature already discussed, derives from particular disciplinary understandings of method, and specific commitments to the nature of knowledge. LA is promoted in a way that positions 'decision making' and 'action' as taking place after the results of data processing precisely because they are assumed to occur without the subjective influence of choice or interpretation (Boellstorff and Maurer 2015). This is the 'mythology' tied up with the technical and analytic capacities of big data (boyd and Crawford 2012), and as we have seen in the literature discussed, the advocacy of LA often seems to promise a closer connection with the 'truth' of educational activity. boyd and Crawford warn against uncritical responses to 'the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy' (2012, 663). LA often

inherits this commitment, sometimes directly translating Anderson's well-known claim that '[w]ith enough data, the numbers speak for themselves' (2008) into a specifically educational declaration: 'let the learners speak for themselves' (Charlton 2013). In other words, the 'black boxing' of LA tends to take place because the methods are often presumed to provide more objective or truthful results.

Key to understanding how LA is able to make powerful claims of truth is recognising the well-known difference between correlation and causation. Mayer-Schonberger and Cukier describe succinctly how correlation functions:

if *A* often takes place together with *B*, we need to watch out for *B* to predict that *A* will happen. Using *B* as a proxy helps us capture what is probably taking place with *A*, even if we can't measure or observe *A* directly. Importantly, it also helps us predict what may happen to *A* in the future. (2013, 53)

This reveals another kind of distancing and abstraction in LA, one that happens at the micro-level of data capture. The underlying premise of the LA field is that data captured from human behaviour can be used to identify, or predict, learning. Rogers critiques this uncritical adoption of 'scientific empiricism', which leads to the questionable identification of causes. As Rogers insightfully shows, the conditions we find in education cannot be understood as 'sufficient causes' for the 'logically deductive explanations and predictions' (2015, 224) required in scientific empiricism. This correlation is precisely the kind of surface observation that LA focusses on, steered by what is found in the data rather than the things that might trigger them.

Importantly however, the tendency in computational analysis is not to deny the difference between correlation and causation, but rather to claim a diminishing need for the latter. Mayer-Schonberger and Cukier admit that 'correlations show *what*, not *why*', but also go on to claim that 'knowing *what*

is often good enough' (2013, 59). This explains the emphasis on a post-analytic 'intervention' and 'action' discussed previously, which reduces the activity of the teacher or student to a *response*. However, the question specific to the context of education is whether knowing 'what' is enough. Although other factors in education – assessment being a prominent example – are clearly premised upon, and concerned with the rigours of causality, LA appears to be emerging as an area that overlooks cause in favour of correlation. This surface understanding underpins the dominant trend of reporting on student progress, and what is perhaps the most well-known example of LA: *Course Signals* from Purdue University³. Such analytics inform the teacher or student 'what' the relationship is between an individual's progress and the course aims, however precisely 'why' those variables align (in the case of good progress) or not (in the case of failure) may not be known. This may be particularly acute in institutions or contexts where contact between students and teachers is reduced. In this way, LA tends to reconstruct roles in education: teachers are encouraged to react to data without necessarily understanding the reasons for doing so, and students are urged to foster uncritical relationships with data reporting, and to understand their educational progress in narrowly defined ways.

The danger of assuming that cause is increasingly irrelevant in the face of evermore attractive correlations is manifold. Principally however, it should be acknowledged that: a proxy necessarily produces something that 'appears' to be what it represents; and correlations are used because they are the things easily *found*, not the things that one actually wants to measure. LA is largely underpinned by behaviourism - a theory of learning which posits that observable behaviour is indicative of learning – which manifests as rather orthodox in contemporary educational research.

As Rogers contends:

the failure of behaviourism was the impossibility of identifying actions through the description of physical behaviour (such as mouse clicks) because the

identification of the meaning of any action requires the comprehension of the context and of the relevant student and educator intentions and their social relations. (2015, 225)

Nevertheless, most examples of LA appear to assume that measurable behaviour (usually interactions with educational software) can directly represent 'learning', or the lack of it. The disciplinary influences of the computer sciences are also important here: in just the same way as the Turing test measures the ability of a computer to perform *as if* it is conscious (Penrose 1990), LA, in effect, only detects activities that *appear* to demonstrate learning. Given the burgeoning power over institutionalised education that LA is beginning to achieve, serious questions must be asked about how seriously we accept, and act upon, the results.

Largely operating with data gleaned from student information systems and LMSs (Siemens 2013), LA is necessarily limited in the kinds of proxies it can work with. The emphasis often shifts to that which is easily captured, rather than that which might represent the complex activities of learning activity. Potentially 'meaningful, relational and unquantifiable social objects such as reasons, norms and rules' (Rogers 2015, 225) are overlooked in LA, precisely because they are often too difficult to measure. As boyd and Crawford suggest, big data isn't simply a means to capture reality in better, more efficient ways, it 'stakes out new terrains of objects, methods of knowing, and definitions of social life' (2012, 665). In other words, the identification of proxies for learning in the form of data has the effect of producing new educational realities. When understood to be authentically representative, correlative data from LMS software construct a narrow vision of education as success or failure (as in the Course Signals example above). This is based on simplistic measurements of behaviour, but also unmistakably 'driven by the demands and world view of managers and the economic framing of education' (Clow 2013, p693).

However, it is not necessarily 'more data' that would provide a solution. Campbell has questioned the ability of data collection and representation to correspond authentically with the complexities of reality, of which education maybe a particularly salient example (2012). For Beer *et al.* simple causes cannot be extracted from such complex systems (2012), calling into question the behaviourist commitments that underpin the LA field. Rogers similarly questions whether 'the social world can be adequately understood by reference to discrete, constant and quantifiable entities' (2015, 225). However, the critical realist approach that Rogers advocates still remains committed to 'the ever deeper uncovering of casual mechanisms that explain action' (Rogers 2015, 229). Methodological diversity may be one way forward for the current educational obsession with data, where the 'deep' data offered by anthropological approaches might support the big data produced through computation (Boellstorff and Maurer 2015).

Predilections for prediction

LA has moved 'from a focus on hindsight to foresight' (Johnson *et al.* 2016, 38), and predictive modelling has emerged as the most mature and widest deployed approach (Clow 2013, Brown 2012). The power and authority achieved by computational data analysis, despite the questionable relation to causes, derives principally from the emphasis on prediction. Once again, the context of this power appears explicit. As Willis suggests, '[t]his is no mere abstraction - student-level data is deployed across modeling to change the future, whether for the intent of producing successful graduates and/or to generate profits' (2014). Prediction in LA is commonly employed in 'estimating how likely it is that individual students will complete a course, and using those estimates to target support to students to improve the completion rate' (Clow 2013, 686). *Course Signals* does precisely this, displaying: 'green, denoting a high chance of success; yellow, denoting potential problems; or red, denoting a high chance of failure' (Clow 2013, 687). Nevertheless, prediction itself entails three

significant problems for education: its reliance on 'the past'; the 'justice' of forecasting; and the kind of future it encourages.

As Clow describes, predictive LA works by comparing current student data, including profile information and statistics from interactions with course materials, with large cumulative sets of these statistics from previous cohorts (2013). The 'big data' associated with recent MOOCs have been suggested to exemplify this approach, using cumulative errors from past groups to guide present students, or to understand the efficacy of video resources and forum posts (Mayer-Schonberger and Cukier 2013). This highlights the problems of sampling, which are often overlooked in large data sets (Halford 2014). The assumption of 'n=all' (Halford 2014) has particular resonance for education in this context, where particular student cohorts can be assumed to substitute for entire populations. Understanding a universal process of learning (if such a thing exists) may not come from analysing the behaviour of a relatively small number of students, at institutions privileged enough to deploy the latest LMS software. Rather, such approaches may establish skewed, but nevertheless powerful, understandings of learning, based on particular technology-rich examples.

Willis frames this general position insightfully along the lines of utilitarian ethics, where the needs of the individual are subordinated to a concern for the many (2014). This raises questions about whether cumulative behaviour can be a useful guide to the conduct of future students. Where the outcome of predictive modelling is nothing other than 'a set of estimated probabilities' (Clow 2013, 687), justification for acting in response becomes unclear. Predictive systems (like Course Signals) preordain human judgement and govern how it is able to function. However, whether accomplished by 'human' decision-making or 'non-human' analytic process, prediction is questionable as a justification for action. Discussing a future scenario of crime prediction using big data, Mayer-Schonberger and Cukier claim that: '[t]o accuse a person of some possible future behaviour is to

negate the very foundation of justice: that one must have done something before we can hold him accountable for it' (2013, 161). They go on to suggest:

The danger is much broader than criminal justice; it covers all areas of society, all instances of human judgement in which big-data predictions are used to decide whether people are culpable for future acts or not. (Mayer-Schonberger and Cukier 2013, 162)

Given that LA is often targeted towards identifying 'at-risk student populations' (Johnson 2016, 38), this issue is particularly significant in education. In predictive analytics, individual students' current behaviour is compared to models derived from the activity of past cohorts. If current behaviour matches a profile of failure, the analytics predict failure, and place the student 'at risk'. While it may be ethically appropriate to act upon students as a result of their actions, to do so based on predictions of *possible actions* may not be.

The final concern with prediction involves the kind of future it attempts to determine: one in which students remain 'on track', and never fail. LA often appears to desire the elimination of mistakes, challenges or interruptions, and to instil instead a deterministic future of success at every stage. Before a student is allowed to 'fail', predictive systems signal 'risk', and encourage the teacher to intervene in order to prevent it. However, this overlooks the educational value of failure (Willis 2014), a part of a learning experience that may lead to deeper engagements, and richer and more profound experiences in the long run.

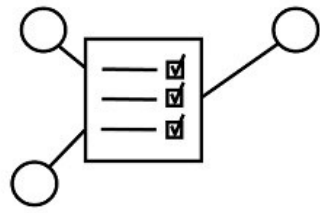
The Learning Analytics Report Card

The Learning Analytics Report Card (LARC) was a small-scale pilot project undertaken at the Centre for Research in Digital Education at the University of Edinburgh. This section will outline the project, not with the intention of providing an exhaustive analysis or justification for the LARC, but rather to offer an example of the ways that educational practices can work towards a critical understanding of LA. The central question underpinning the project was: 'How can students and university teaching teams develop critical and participatory approaches to educational data analysis?' This approach was motivated by the sense that LA was 'distanced' from the practices of teaching and learning, as described previously in terms of disciplinary barriers and hidden processes.

The primary aim of the project was not to produce precisely functioning algorithmic processes, but rather to encourage students (and teachers) to critically interrogate the production of LA data, and to consider what kind of automated processes might be involved in educational decision-making. The pilot project was therefore embedded within three course options from the MSc in Digital Education programme at the School of Education; a fully-online programme offering a range of courses at post-graduate level⁴. Rather than operating independently from the courses in question, the LARC was aligned with already existing course themes (including learning theory, online presence, digital research methods, and privacy and surveillance), and introduced as a provocation for critical reflection on the contemporary condition of 'digital' and 'distanced' higher education. Piloted with this specific group – who were invested in a digitally-mediated relationship with the university as well as being exposed to critical perspectives on the experience - the project sought to promote an understanding of data analysis that is not only in need of interpretation, but also increasingly co-constitutive of social life, and in particular the slice of social life that pertains to educational activity. Therefore, the focus of this section of the paper is not the technical functions of the LARC⁵, but rather on an analysis of the ways it might illustrate critical and interventionist teaching practices in the area of LA.

Student involvement in the project was considered vital and incorporated at all stages of the project. The design, development, and testing phases of the LARC were informed by formal student representation, and this was motivated by a general concern for ethical practices in data collection. As Johnson *et al.* contend ‘many institutions view privacy concerns and the safety of student data as obstacles’ (2016, 39). This project sought to build on established work that has taken the opposite approach, such that from the Open University in the UK, which has established important student-led policy development on the ethical use of data⁶. However, the LARC was not intended to ‘solve’ the dilemmas of educational data analytics through the straightforward production of a ‘student-centred’ technology. Rather, the aim was to highlight the entangled and complicit conditions of data capture and surveillance, and to offer ways for students to both experience and reflect upon forms of participation within the LA process. Neither was it expected that the LARC would ‘solve’ the problems of ‘blackboxing’, rather the purpose was to foreground the ways different aspects of the analytics process are made visible, while others are concealed, and highlight the inherent inscrutability of algorithmic operations (Ziewitz 2015).

The LARC functioned by capturing student’s course-related activity data within an institutional LMS (Moodle), and presenting a summary of an individual’s academic progress in the form of a textual report. However, rather than manifesting exclusively through hidden and inaccessible institutional data aggregation and analysis, the LARC offered students an opportunity to ‘play’ with their data; to choose what is included or excluded, and when the report is generated – hence the acronym ‘LARC’. This facility for choice, although limited, was intended to explore the capacity for participation while still retaining the ‘reporting’ function of LA. The LARC interface consisted of a web form, accessible only through the institutionally protected login, which provided various options for generating the report (see fig 1).



Choose what you'd like the report to cover:

Course

Week Commencing

- Attendance
- Engagement
- Social Interaction
- Performance
- Personal

Ok, report on me!

[View my archived reports](#) | [LARC Project Site](#)

Figure 1: The LARC web form interface

Drop down boxes at the top of the interface allowed users to select a course from which to generate the report, and importantly, the specific week from which to view activity (see fig 1). This choice of *when* to generate the report was considered a crucial part of learner agency within the project, and students were encouraged to take charge of the process and activate the LARC whenever they desired.

Student choice related to the content of the report, and thus the kind of data that the LARC would capture and analyse, was facilitated through the selection of report themes: 'attendance', 'engagement', 'social interaction', 'performance', and 'personal' (see fig 2). These were intended to suggest different perspectives on course activity, and allowed students to select one or more to generate the report, and thus tailor their subsequent result towards particular aspects of their

course interactions. Basic data was gathered from specific areas of the LMS for each report theme: 'attendance' from date and time of login; 'engagement' from frequency of interaction with course resources; 'social interaction' from discussion forum module data; 'performance' from aggregating the other theme data and comparing it to the student group; and 'personal' from interactions with profile information and introductory course tasks.

<input type="checkbox"/> Attendance	<input checked="" type="checkbox"/> Performance
<input checked="" type="checkbox"/> Engagement	<input type="checkbox"/> Personal
<input type="checkbox"/> Social Interaction	

Ok, report on me!

Figure 2: a subsection of the LARC web form interface, showing the choices available for each report.

Algorithmic calculations and data-to-text processes were developed to present this data in the form of a 'report card', consisting of a number of automatically generated sentences that corresponded to the report themes selected by the student (see figs 3 and 4). An outline of the calculations for the 'social' category is presented below as an example:

Depending on the number of times the student accessed the course discussion forum within the selected week, the following thresholds were used to determine the resulting sentence in the report:

- < 4 "not very social"
- < 8 "somewhat social"
- < 12 "very social"
- >= 12 "extremely social"

For this pilot project, these thresholds were set by the research team, with the recognition that faculty would need to be involved in these decisions in future development of the LARC, in order to ensure that thresholds accurately represented their perception of behaviour ranges in particular course contexts. These thresholds were not presented to the students making use of the report card, although a later edition of the LARC included some data related to student activity. This was principally a pedagogical decision intended to highlight the difficulty in understanding algorithms, not only technically, but also in their broader social influence. Simply revealing the technical workings of the underlying algorithms would not have reflected current educational conditions in which data mining practices are largely non-transparent. The the aim was to highlight the deeply embedded condition of algorithms, and the idea that coming to ‘know’ them can only ever be partial and limited (Ziewitz 2015).

Your Report:

Your attendance has in general been good. Maintaining consistent attendance will allow you to engage consistently with the course content and build your understanding of the core concepts.

Attendance is key to achieving your aims on the course, and you are giving yourself this opportunity by attending regularly.

You have usually been extremely social during the course.

You are in the lowest third for social interaction, but the middle third for attendance.

Leave a comment

Figure 3: An example anonymised report from the LARC, showing a number of sentences produced from selected report themes.

Your Report:

Your attendance has in general been excellent. Attendance often relates directly to performance, so try to maintain your current level.

You have mostly been extremely engaged with the course content. Continuing to read the course materials will allow you to expand your own aims and objectives in relation to those of the course.

You have usually been extremely social during the course but in week 12 you interacted less with others.

Leave a comment


 

Figure 4: An example anonymised report from the LARC, showing a number of sentences produced from selected report themes.

A facility for student to add comments was provided underneath each report (see figs 3 and 4), and was intended to expand the scope for interactivity in the LARC. Comments encouraged students to reflect, not only on their activity within the course, but also to critically discuss the efficacy of the generated report, the kind of language used, and experience of participating in decision-making (and the lack of it) concerning the analytics process itself. Through facilitating choice in the report themes, learners had the ability to reflect on what an ‘active’, ‘engaged’ or ‘valued’ student might be, and to recognise how the automated algorithmic processes of data analytics might interpret their online activity, and make decisions behind the scenes to which they were not party. All generated reports were archived and available to individual students, and any report could be downloaded and shared if desired. The generation of a report on previous activity, combined with the ability to comment and reflect on the result, was also intended to offer a deliberate alternative to predictive LA that assign ‘risk’ to student conduct. Rather than using predicted behaviour to encourage intervention, the LARC was thus designed to foreground prior conduct, and to motivate

students to reflect on ways that they might act to alter or maintain the performance implied in their reports.

Student Feedback

Within the limited scope of this pilot project it would be difficult to draw definitive conclusions about the LARC's capacity to foster critical awareness amongst participants. Nevertheless, the feedback from students offers some insight into the ways students were able to reflect on the experience of 'participating' in LA, in the ways described above. Over a 12-week period, the 12 students who were given access to the LARC generated 208 reports. 26 reports were accompanied with comments, indicating a partial engagement with this feature, however, feedback was also given within the course discussion spaces, as a result of embedding the LARC within course activities, as discussed previously.

On the whole, comments and discussions tended to be concerned with the accuracy of the analytics, and the experience of receiving automated feedback. This was perhaps due to the novelty of the activity, as expressed by one participant: 'to have "computerised" (through human generated algorithms of course) feedback is a newish experience!' Nevertheless, an awareness of the entanglement of social and technical concerns is expressed here. Critical views of the capacity for the system to acknowledge the full range of learning activities was also expressed. One participant commented:

it cannot account for the time I might have been engaging informally with the course (thinking about my dissertation topic, dipping into the course textbook, the level of interest demonstrated in the first assignment).

This highlights the constrained view of data analytics in the context of learning, suggested by this student to encompass a range of spaces and times outside of direct interaction with class activities.

This is reflected in a comment from another participant:

it is possible for a student to be not fully engaged with the course content but still engaged with reading material from other related sources... How would the tutor know this from analytics only?

This underscores the problems of assuming that easily accessible data is fully representative of phenomena being studied (boyd and Crawford 2012), and also by implication, the way such accessibility of data reshapes the research agenda in education (Eynon 2013). Reading around a topic becomes less important than registering contact with the institutional LMS. This also reflects the discussion of the narrow disciplinary fields underpinning LA, and points to the need for broader understandings of the range of spaces and activities which constitute learning in interaction with the institution. However, crucially, these discussions centred on the representative flaws of the data. It was quickly recognised that the analytics were 'superficial', as one participant commented, and was concerned with ranking and grouping students, rather than seeking to 'understand' the character of activity. This reflects the process of 'metadatification' suggested by Andrejevic *et al.* (2015), in which content is overlooked in place of information about it. Discussing the example of Google, they suggest that this focus on metadata is utilised to 'arrange and sort people and their interactions' (2015, 381) according to predetermined aspects of the underlying algorithm. Elsewhere, another participant expressed distrust of the reports, commenting: 'I need to observe the results over a period of time before I can gain more confidence in the information it's providing.' This highlights the importance of relationship building in educational activity, and calls into question the transactional functions of automated systems in this context. Represented in such narrow terms, students become enfolded in the instrumentalist logics of LA, which seek the most efficient

educational relationships where metadata is used to reduce student interaction with teachers to calculated responses.

Occasionally, some of this broader discussion of the cultural implications of LA surfaced, such as in the comment: 'should students be made aware of how analytics are functioning and how they might feel as though they are being "watched". Or does this only add to the creation of a surveillance culture?' This points to concerns for ethics in educational data analytics, and questions about which aspects of student learning should remain private (Eynon 2013). A sense of unease at the pervasive and agential aspects of analytics were also apparent, as in a comment from another participant: 'Well it is interesting that HAL has crept into my life.' This popular culture reference to the dystopic potentials of artificial intelligence highlights, not only a tendency to personify non-human algorithmic agencies, but also perhaps a concern for the loss of human decision-making and interaction in the context of the educational institution. While these discussions may demonstrate some productive engagement with the 'opening up' of particular LA processes, more work could be done with the LARC to foster teacher and student discussion about the broader social, political and economic drivers embedded in such systems. This is future teaching work which could be carried out with software such as the LARC.

Conclusions

The emerging field of LA involves a 'blackboxing' of underlying algorithmic processes that obscures inner-workings for the people often tasked with responding to the results: teachers and learners. As Latour highlights, the more successful a technology becomes, the less its internal workings are placed under scrutiny (1999) and this is an area for LA to develop as its prominence and power increase. However, the concealment of computational processing in education is more than simply technical; it is the impetus to look *only* at the data output, rather than the activities of education

that surround it, and the broader social, political and economic conditions in which it is embedded. The 'pressure towards performance management, metrics and quantification', as highlighted by Clow (2013, 685), is not visible in the data itself. However, if teachers and learners are encouraged to adopt critical and questioning perspectives on data reporting, this may be one way to resist managerialist pressures and foster knowledge cultures that view the university within broader contexts of power.

As we have seen, big-data is based on correlations not causality (Mayer-Schonberger and Cukier 2013). The dilemmas of this distinction are not new, and the allures of big data should not lead to 'making the same old statistical mistakes on a grander scale than ever' (Halford 2014). For Mayer-Schonberger and Cukier, '[c]orrelations let us analyze a phenomenon not by shedding light on its inner workings but by identifying a useful proxy for it' (2013, 53), yet 'inner workings' may be the most valuable aspects to examine in education. We may need to 'understand the meaning of actions rather than the cataloguing of behavioural markers' (Rogers 2015, 229). As boyd and Crawford warn, 'taken out of context, data lose meaning and value' (2012, 670), and education needs to balance the current fixation on data with other methods that are able to interrogate situations and circumstances in ways that provide additional, and deeper contextual insights.

Critical attention should also be given to the role of prediction in education, and the powerful way it conditions our agency and autonomy in the learning process. As Mayer-Schonberger and Cukier insightfully explain, '[w]ere perfect predictions possible, they would deny human volition, our ability to live our lives freely. Also, ironically, by depriving us of choice they would exculpate us from any responsibility' (2013, 161). In other words, if LA prediction actually worked, we would have to exonerate students, teachers and institutions from any responsibility for educational success, or lack of it. For Willis '[l]earning analytics must respond to such capacity to rewrite a non-deterministic

future' (2014), and this may indeed be most valuable to the learning process: accommodating the potential richness of interruption, chance, and failure.

The LARC pilot project was outlined in this paper as one example of developing critical practices around computational data analysis in education. While limited in scope, this project has sought to engage students as active agents of their data-driven assessment, and involve them in the critical questions faced by universities in the digital age. The LARC project has attempted a critical *and* practical engagement with LA, encouraging the convergence of interdisciplinary approaches, and the involvement of multiple educational stakeholders. In this sense, data analysis has been explored in a way that is itself *educational*, opening processes for discussion and community exchange, developing critical understanding of computation, as well as ways of knowing through data, rather than simply *acting upon* education from a distance. One of the central critical issues raised through the LARC was the sense that control, agency, and power are never uncomplicated where data processing is involved, and that any 'student-centred' LA could never be entirely devoid of institutional influence and authority. Rather, the underlying aim of the LARC was to provoke awareness, and encourage students to reflect on the experience of a situation in which agency was not straightforward, and in which 'choice' was accompanied with concealment and algorithmic decision-making. The pilot has also revealed a crucial need for further engagement with the cultural studies of data in education. Future critical work with LA needs to address the histories of the technologies employed in data capture, as well as the related political and economic conditions which increasingly retain the efficiencies of data computation, and in so doing continue to restructure education and its subjects.

References

Anderson, Chris. 2008. "The End of Theory: The Data Deluge Makes the Scientific Method Obsolete". *Wired Magazine*. Accessed February 1st, 2011, http://archive.wired.com/science/discoveries/magazine/16-07/pb_theory

Andrejevic, Mark, Hearn, Alison and Kennedy, Helen. 2015. "Cultural studies of data mining: Introduction". *European Journal of Cultural Studies*. 18 (4-5). pp.379-394

Andrejevic, Mark and Gates, Kelly. 2014. "Editorial. Big Data Surveillance: Introduction". *Surveillance & Society* 12(2). pp. 185-196.

Barneveld, Angela van, Arnold, Kimberly E. and Campbell, John P. 2012. "Analytics in Higher Education: Establishing a Common Language", *ELI Paper*, vol. 2012, no. 1, pp. 1-11, 2012.

Beer, Colin, Jones, David, Clark, Damien. 2012. "Analytics and complexity: learning and leading for the future". In: *29th Annual Conference of the Australasian Society for Computers in Learning in Tertiary Education (ASCILITE 2012): Future Challenges, Sustainable Futures, 25-28 Nov 2012*, Wellington, New Zealand.

Beer, David. 2009. "Power through the algorithm? Participatory web cultures and the technological unconscious". *New Media & Society* 11(6): pp.985–1002. DOI: 10.1177/1461444809336551

Boellstorff, Tom. and Maurer, Bill. 2015. "Introduction". In T. Boellstorff and B. Maurer (eds), *Data, Now Bigger and Better!* Chicago: Prickly Paradigm Press

boyd, d., & Crawford, K. (2012). "Critical Questions for Big Data". *Information, Communication & Society*, 15(5), 662–679.

Brown, Malcolm. 2012. "Learning Analytics: Moving from Concept to Practice". *Educause Learning Initiative*. Available at: <http://net.educause.edu/ir/library/pdf/elib1203.pdf>

Campbell, Gardner. 2012. "LAK12 conference presentation". *Canada: Learning Analytics and Knowledge 2012*

Campbell, John P., De Blois, Peter B. and Oblinger, Diana G. 2007. "Academic Analytics: A New Tool for a New Era". *Educause Review*, 42 (4): 40-57.

Camps-Febrer, Franc. 2012. "How Learning Analytics Can Empower Teachers and Encourage Critical Thought". *The Knewton Blog*. Available at: <https://www.knewton.com/resources/blog/teacher-tools/how-learning-analytics-can-empower-teachers-and-encourage-critical-thought/>

Charlton, Patricia. 2013. "Let the learners speak for themselves: the potential of learning analytics and big data event". Available: <https://www.lkdev.ioe.ac.uk/lklinnovation/big-data-and-learning-analytics/>

Clow, Doug. 2013. "An overview of learning analytics". *Teaching in Higher Education*, 18(6), pp.683–695. Available at: <http://www.tandfonline.com/doi/abs/10.1080/13562517.2013.827653>.

Cooper, Adam. 2012. "What is Analytics ? Definition and Essential Characteristics." *CETIS Analytics Series*, 1(5), pp.1–10.

ECAR-ANALYTICS Working Group. 2015. "The Predictive Learning Analytics Revolution: Leveraging Learning Data for Student Success". *ECAR working group paper*. Louisville, CO: ECAR, October 7, 2015

Eynon, Rebecca. 2013. "The rise of Big Data: what does it mean for education, technology, and media research?" *Learning, Media and Technology*, (March), 1–4.

<http://doi.org/10.1080/17439884.2013.771783>

Gasevic, Dragan., Dawson, Shane., Siemens, George. 2015. "Let's not forget Learning Analytics are about Learning". *Tech Trends*. 59(1). pp. 64-71.

Gillespie, Tarleton. 2014. "The Relevance of Algorithms". In *Media Technologies*, ed. Tarleton Gillespie, Pablo Boczkowski, and Kirsten Foot. Cambridge, MA: MIT Press. pp. 167-194.

Harford, Tim. 2014. "Big data: are we making a big mistake?" *FT Magazine*. March 28. Available at:

<http://www.ft.com/cms/s/2/21a6e7d8-b479-11e3-a09a-00144feabdc0.html#axzz2xOGR9R7f>

Johnson, Larry., Adams Becker, Samantha., Estrada, V., Freeman, A. 2014. NMC Horizon Report: 2014 Higher Education Edition. Austin, Texas: The New Media Consortium. Available at:

<https://www.nmc.org/publication/nmc-horizon-report-2014-higher-education-edition/>

Johnson, Larry., Adams Becker, Samantha., Estrada, V., and Freeman, A. 2015. NMC Horizon Report: 2015 Higher Education Edition. Austin, Texas: The New Media Consortium. Available at:

<http://cdn.nmc.org/media/2015-nmc-horizon-report-HE-EN.pdf>

Johnson, Larry., Adams Becker, Samantha., Cummins, M., Estrada, V., Freeman, A., and Hall, C. 2016. NMC Horizon Report: 2016 Higher Education Edition. Austin, Texas: The New Media Consortium. Available at: <http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf>

Latour, Bruno. 1999. *Pandora's hope: essays on the reality of science studies*. Cambridge, Massachusetts: Harvard University Press.

Mayer-Schonberger, Viktor., and Cukier, Kenneth. 2013. *Big Data: A revolution that will transform how we live, work and think*. London: John Murray.

McKenney, Susan. and Mor, Yishay. 2015. "Supporting teachers in data-informed educational design". *British Journal of Educational Technology*. 46(2). pp.265–279 doi:10.1111/bjet.12262

Norris, Donald M. and Baer, Linda L. 2013. "Building Organizational Capacity for Analytics". *Educause Review*

Penrose, Roger. 1990. *The Emperor's New Mind: concerning computers, minds, and the laws of physics*. London: Vintage.

Perrotta, Carlo. & Williamson, Ben. 2016. "The social life of Learning Analytics: cluster analysis and the 'performance' of algorithmic education", *Learning, Media and Technology*, DOI: 10.1080/17439884.2016.1182927

Rogers, Tim. 2015. "Critical realism and learning analytics research: epistemological implications of an ontological foundation". *Proceedings of the Fifth International Conference on Learning Analytics*

And Knowledge - LAK '15, pp.223–230. Available at:
<http://dl.acm.org/citation.cfm?id=2723576.2723631>.

Siemens, George. 2013. “Learning Analytics: The Emergence of a Discipline”. *American Behavioral Scientist*, 57(10), pp.1380–1400.

Striphas, Ted. 2015. “Algorithmic Culture”. *European Journal of Cultural Studies* 18(4-5)

Tufekci, Zeynep. 2015. “Algorithmic Harms beyond Facebook and Google: Emergent Challenges of Computational Agency” *Colorado Technology Law Journal*. 13(2).

Williamson, Ben. (ed.) 2015. *Coding/Learning: Software and digital data in education*. Stirling: University of Stirling. Available:
https://codeactsineducation.files.wordpress.com/2015/08/coding_learning_-_software_and_digital_data_in_education.pdf

Willis III, James E. 2014. “Learning Analytics and Ethics: A Framework beyond Utilitarianism”. *Educause review*. Available at: <http://er.educause.edu/articles/2014/8/learning-analytics-and-ethics-a-framework-beyond-utilitarianism>

Ziewitz, Malte. 2015. “Governing Algorithms: Myth, Mess, and Methods”. *Science, Technology, & Human Values* pp.1-14. DOI: 10.1177/0162243915608948

¹ See <http://nextgenlearning.org/topics/learning-analytics>

² For example: <http://www.laceproject.eu/lace/>

³ <http://www.purdue.edu/uns/x/2009b/090827ArnoldSignals.html>

⁴ <http://online.education.ed.ac.uk/>

⁵ Isard, A. and Knox, J. 2016. Automatic Generation of Student Report Cards. 9th International Natural Language Generation conference. Edinburgh, Scotland, September 5-8, 2016. Available at: <http://www.macs.hw.ac.uk/InteractionLab/INLG2016/proceedings/pdf/INLG33.pdf>

⁶ <http://www.open.ac.uk/students/charter/essential-documents/ethical-use-student-data-learning-analytics-policy>