Datafication of education: a critical approach to emerging analytics technologies and practices

Ben Williamson, University of Edinburgh

[Author version of chapter published in Beetham, H. & Sharpe, R. (eds) 2019. *Rethinking Pedagogy for a Digital Age*, 3rd ed. London: Routledge]

The availability of 'data' through digital information systems has become a defining topic and problem of recent times. For companies, access to digital data is a source of 'business intelligence' used to make efficiency savings and gather profit. Governments treat data as insight into people's behaviours and wider social trends to inform policymaking. Data are also used for more controversial purposes. Recently, Facebook user data was exploited by the data analytics consultancy Cambridge Analytica to 'micro-target' political advertising to voters in the 2016 EU referendum and US election, while data breaches have become a common occurrence. As political, commercial and public awareness has grown around data controversies, researchers have begun to develop better understanding of the consequences and ethics of data, analytics, algorithms and artificial intelligence (AI Now 2018).

Within the field of education and pedagogy, however, data remain relatively under-researched, although significant conceptual and technical development in learning analytics, adaptive learning software, and artificial intelligence has emerged from academia, industry, consultancies, think tanks, and government sources alike (Williamson 2017). Data can be used as insight into processes of learning, the effects of pedagogies, curriculum design, and learning gain over time, especially as researchers in 'education data science' develop increasingly fine-tuned technologies and methods (Cope & Kalantzis 2015; Lang et al 2017; Piety, Hickey & Bishop 2014). Moreover, data may be used to predict outcomes, detect risks, and to 'personalize' the education system around individuals' needs (Bulger 2016). A lucrative industry in data-driven educational technologies has thrived on claims that data signify a shift from standardized tests to adaptive, 'real-time' assessment technologies, and from school census data to individualized tracking and profiling (Boninger & Molnar 2016). Such data-centred processes are affecting the early years of education, schooling, and higher education alike, often with unintended and perverse consequences (Bradbury & Roberts-Holmes 2018; Manolev, Sullivan

& Slee 2018; Roberts-Mahoney, Means & Garrison 2016; Selwyn 2015; Williamson 2018).

This chapter offers a critical introduction to the 'datafication' of education to encourage better understanding and debate about data in relation to pedagogy and curriculum, in particular by drawing attention to the history, epistemology, social consequences, cultural contexts, politics and ethics of datafication. Datafication can be understood as ways of seeing, understanding and engaging with the world through digital data (Gray 2016). However, this definition of datafication glosses over a number of complexities. The rest of this chapter offers definitional clarity, applied to some particular issues and problems in education, to help us make better sense of the consequences of datafication in education, and therefore inform better decisions about learning, teaching, and curriculum design.

Data history

Datafication has a long history, going back at least as far as efforts during the industrial revolution to capture statistical knowledge of the state, society and its population, and then to use that knowledge to come up with better institutions and practices of management and intervention (Ambrose 2016; Beer 2016). State power has long been tied to statistical practices such as archiving, census-taking, indexing, cataloguing, and record-keeping. By turning people into numbers in an increasingly statistical society, it became possible for governments, bureaucracies and public agencies to sort them into categories and population segments, to derive 'norms' and 'regularities' from the aggregated numbers and then evaluate and judge people against them. 'This is power through numerical knowledge of the people being governed, their properties, and the patterns of social life' (Beer 2016: 44), and allows individuals and social groups to be identified for intervention.

In terms of education, Michel Foucault (1991: 147) powerfully articulated how schools and classrooms act as 'learning machines' deploying numbers and tabulations for 'supervising, heirarchizing, rewarding' as well as timetabling, grouping and controlling students. The development of school management systems at the end of the twentieth century created new kinds of digital learning machines for capturing, calculating and categorizing information about students. Writing at about the same time, though in the context of higher education, Jean-Francois Lyotard (1984: 4) noted that the 'miniaturization and commercialization of machines is already changing the way in which learning is acquired, classified, made available and exploited.' He anticipated any knowledge not translatable into ‘quantities of information’ would eventually be abandoned, while only knowledge ‘produced in order to be sold’ and consumed would continue to hold ‘value’ (Lyotard 1984: 4). Lyotard presciently foresaw that digital information would become a highly valuable commodity for learning in increasingly computerized societies, as many commercial producers of online courses have since exploited. The datafication of education, then, is part of a series of historical developments in statistics, state power, quantification, computation, and valuation culminating with the expansion and intensification of digital information systems and ‘big data’.

Although there are clear continuities from the past to the present, the current version of datafication through big data also represents a rupture with the past. Early historical developments in the datafication of education have already evolved into the global data-driven performance comparisons of international large scale assessments (LSAs) such the OECD’s PISA tests (Sellar, Thompson & Rutkowski 2017). The assessment data that dominates LSAs is sampled, collected at long periodic intervals, and slow to analyze. New digital datafication technologies such as ‘learning analytics,’ by contrast, harvest data in real-time as students complete tasks, enable high-speed automated analysis and feedback or adaptivity, and can capture data from all participants rather than a sample (Lang et al 2017). They also allow individuals to be compared against each other and with aggregated norms calculated in massive datasets, rather than the broad-brush comparison of national systems enabled by LSAs. Nonetheless, with the emphasis on quantification, the construction of norms, and methods of comparison, big data techniques such as learning analytics importantly need to be understood as part of a long history of measurement in education (Gorur, Sellar & Steiner-Khamsi 2019). Just as LSAs have standardized schooling around the demands of the tests, the danger of using big data to drive curriculum design is that teaching and learning methods become confined to those that produce measurable outcomes such as credentialized achievements or metrics of individual progress.

Technicalities of data

In technical terms, datafication is a process of transforming diverse processes, qualities, actions and phenomena into forms that are machine-readable by digital technologies (Kitchin 2014). Datafication allows things, relationships, events, and processes to be examined for patterns and insights, often today using technical processes such as data analytics and machine learning. These data-analysing systems rely on complex algorithms to join up and make sense out of thousands or millions of individual data points. As a technical concept, big data has attained popular currency in recent years, but not all digital data are big data. Rob Kitchin (2014: 28) has usefully contrasted big data with 'small data.' By small data he means collections of data that are sampled, slow to collect, and of relatively limited variety and volume. Even large-scale assessment data would fall into the category of small data by this definition. By contrast, big data is very large in volume, exhaustive in terms of its capture of whole populations rather than samples, is collected at speed and continuously, has very high variety, and enables very fine-grained identification. However, as Kitchin (2014) also notes, significant current efforts are underway to 'scale up' small data into larger datasets, making them reusable and linking them with other data to make them amenable to analysis. So when we talk of datafication we are usually talking about different kinds of data, although the contemporary promise is that scaled-up, linked or big data can be analyzed to generate intelligence and insight that was previously unavailable.

The technical and informational qualities of datafication in education are significant because they determine how students, teachers, schools, universities, and whole systems can be measured, inspected, evaluated, judged and so on (Anagnostopoulos et al 2013). The technicalities of an international large-scale computer assessment, for example, consist of software loaded on computer hardware in schools; network connections that allow the student responses to flow to a central repository; spreadsheets for tabulating the results as data; servers and storage facilities for holding the data; security and encryption services to keep it safe; analytic software for processing the results; visualization programs for communicating it, and so on. At each stage, the technicalities of datafication influence what data it is possible to collect and examine, and thus what results may be generated.

In UK higher education current efforts are being made to link data captured from institutions' learning management systems, electronic reading lists, and learning analytics platforms to other large-scale governmental and administrative datasets, largely through new infrastructural arrangements to make data interoperable across different platforms (Williamson 2018). Connecting individual-level learning data to other longitudinal data sources opens up new analytical and interpretive possibilities to isolate courses and institutions that contribute to boosting students' 'learning gain' or to later graduate outcomes, careers, and earnings. The potential here is to empower HE institutions and students with insights into the courses and providers that perform best in terms of measurable learning progress and preparing students for high-income graduate roles. Students can make choices based on such data, while providers can use the data to target under-performing areas of the curriculum for improvement, or to change the pedagogy to make it more engaging for students. Politically, it also, however, allows institutions and courses to be rated and ranked on metrics such as graduate earnings, thereby shifting the priorities of the sector to focus on career-readiness and the demands of labour markets. As such, the data infrastructure and technicalities of both big and linked data in HE are making possible analyses that are producing intelligence on university performance. The interoperable digital system is not merely technical, but a network of political technologies to shape how HE is perceived and practised.

Data epistemologies

Thinking epistemologically about datafication concerns what we can *know* from data. Datafication rests on the assumption that patterns and relationships contained within datasets inherently produce meaningful, objective and insightful knowledge about complex phenomena (boyd & Crawford 2012). That is to say, the data captured in a computer *really* represent what is 'out there,' untouched by human interference and independent of the measuring process. From this epistemological standpoint, data are 'raw', impartial, detached, and can be taken at face value as reflecting the truth about the world (Gitelman & Jackson 2013). As Rob Kitchin (2014: 132) has argued, this empiricist epistemology assumes that through the 'application of agnostic data analytics the data can speak for themselves free of human bias or framing.' It has given rise to boosterist claims that ever-bigger datasets offer better, less biased, and more insightful knowledge than has ever been available before.

For critics, however, this empiricist epistemology is flawed because all data are always actively produced, framed, and sampled; data are not simply natural and essential elements that are abstracted from the world in neutral and objective ways to be accepted at face value (Gillespie 2014; Jurgenson 2014). Data may be big, but at the same time, often tend to erase complexities, context, meanings, and causal factors, so producing highly partial and incomplete renderings of reality (Lupton 2015). Different interpretations may also be possible from the same data. In one recent example, 29 teams of data scientists reached different conclusions from their analysis of the same dataset, suggesting that significant variation in the results of analyses of complex data may be difficult to avoid and that subjective analytic choices influence research results (Silberzahn et al 2018).

Within education the empiricist epistemology can be found in the many commercial providers of data analytics services sold to schools and universities. The education company Pearson, for example, has produced several reports on the potential of digital data to reveal new generalizable truths about education, learning and assessment (DiCerbo & Behrens 2014; Hill & Barber 2014). Through vast data collection and analysis efforts, it is presumed, a clearer picture of the realities of learning—even at the scale of the measured individual student—will become available. Those data can then be employed to inform future interventions, to target teaching, and to personalize the learning experience, so changing the 'reality' of the student's experience based on quantitative interpretations.

As such, datafication is not just an epistemological matter, but ontologically significant because it has the potential to bring different version of reality into being (Ruppert 2012). To offer a simple example from the datafication of education, when a child enters a database, she is chopped up into data points, turned into bits, aggregated with other data, evaluated against norms and so on. Over time, as more data becomes available from the student's activity, it becomes possible to generate a data profile of her skills, progress, abilities, and knowledge—often known as a 'student model'—which can be compared with regularities in massive datasets. Sometimes these profiles are called 'data doubles,' as if they represent a digital shadow version of the profiled individual. But, importantly, the data can always be called up and arranged differently—data doubles are really data multiples (Finn 2015). When one of these data multiples gets selected as the student model, it becomes a make-believe substitution which can then be used to inform how the teacher approaches that student, or how an algorithmically personalized learning program assigns her tasks. As such, the substitute profile built out of the data takes an active ontological role in shaping the 'real life' of the student—a process that could always have been done otherwise, with different real world results. The data play a part in 'making up' the student.

Of course, all curricula are addressed to an ideal student that they then call into being. The question here is how the software for data collection is configured to model the student—what categories are preset to sort the data—and how this affects the curriculum choices of the teacher. A learning analytics platform in a university, for example, might be preset to capture data that indicates whether students are developing the required 'graduate attributes'. Several recent data analytics platforms for schools are designed to capture data about students' 'noncognitive' social and emotional learning according to categories such as 'growth mindset' and 'grit', which therefore results in a student model defined in the psychological terms of psychometric classification. In both cases, the graduate with measurable attributes and the student with noncognitive skills is called into being by the categories programmed in the software, with teachers then required to adjust their curricula and pedagogies to ensure students experience growth and development in these categories.

Social data

As already indicated, data do not pre-exist the practices and technologies that bring them into being. As such, data are products of social practices. Datafication is accomplished by social actors, organizations, institutions and practices. In terms of actors, data scientists, data analysts, algorithm designers, analytics engineers and so on have become contemporary experts in the examination of data of all kinds (Gehl 2015). These people or experts are housed in businesses, governments, philanthropies, social media firms, financial institutions, which have their own objectives, business plans, projects and so on, and that frame how and why digital data are captured and processed (Housley 2015). In this sense, datafication can be defined socially because it is always socially situated in specific settings and framed by socially-located viewpoints, despite its advocates' claims to quantitative objectivity and impartiality.

In education, new 'education data scientists' and learning analytics practitioners, engineers and vendors of personalized learning platforms, even entrepreneurs of artificial intelligence in education, are all now bringing their own particular forms of expertise to the examination and understanding of learning processes, teaching practices, schools, universities and educational systems (Cope & Kalantzis 2015; Lang et al 2017; Piety, Bishop & Hickey 2014). Many are supported by funding streams from venture capital firms, philanthropic donations from wealthy technology entrepreneurs, and impact investment programs, which all direct financial resources to the datafication of education and thus shape what objectives and priorities are pursued (Williamson 2017). Commercial companies have become especially significant social actors in education technology and datafication development. Mark Zuckerberg of Facebook, for example, is a generous donor to personalized learning projects and programs through his Chan-Zuckerberg Initiative, recently partnering with the Bill and Melinda Gates Foundation to support new scientific approaches to the psychological and neuroscientific measurement of complex learning processes. These powerful social actors are ultimately seeking to remake public education in ways that reflect the worldviews, technical capacities and business objectives of the Silicon Valley technology sector.

Data also bear effects on teachers as social actors in schools, colleges and universities. With the rise of LSAs, teaching has already become increasingly standardized, with teachers expected to both teach to the tests and to undertake extensive data collection exercises on their own students. New classroom apps such as ClassDojo are extending these data collection efforts into the pedagogic routine of the classroom itself, as teachers are required to input information about students' behaviours according to the categories programmed into the device. With learning analytics and other big data technologies, the machines can be understood to be augmenting the teacher's role, not just by aiding with data collection but by processing those data and making decisions based on algorithmic calculations on the behalf of the teacher. The algorithm becomes a social actor in the classroom, but also changes the social role of the teacher by changing teaching practice into a set of human-machine hybrid tasks where some decisions are made automatically based on calculations from the available data. The subjective human teacher, arguably, becomes partly 'robotized' by deferring to the objective calculations performed by the machine rather than referring to personal professional judgment, experience, and expertise. This means the teacher is socialized to perform professional responsibilities in new ways to support the demands of constant data collection and computational processing.

Moreover, datafication needs to be defined socially because much data is captured from the social world—people, institutions, behaviours and the full range of societal phenomena are the stuff of data. As Geoffrey Bowker (2013: 170) has memorably put it, 'if you are not data, you don't exist'! People are data; societies are the data. Even more consequentially, these social data can be used to reshape social behaviours. Bowker (2013: 168) adds that as data about people are stored in thousands of virtual locations, reworked and processed by algorithms, their 'possibilities for action are being shaped'. Within education institutions, students' identities as individuals and as members of the university are constructed through their access to specific systems—registration, access to the VLE, library catalogue, digital reading list engagement, valued academic content, assessment systems, plagiarism detection software, student records, and learning analytics traces. Understood in this way, technical systems can be understood to bring students into existence as data, which can then be used to infer what student can and cannot do, and what interventions should be taken as a consequence of those calculations.

They also socialize students to expect—and find it normal—to be constantly subject to surveillance and monitoring.

Data power

The new actors undertaking datafication are invested with a certain form of data power. Expert authority, as William Davies (2017) argues, increasingly resides with those who can work with complex data systems to generate analyses, and then narrate the results to the public, the media and policymakers. This is why governments are increasingly interested in capturing the digital traces and datastreams of citizens' activities. Governments are eager to learn from the successes of online platforms and pursuing new models of 'government as collective intelligence' (Mulgan 2016). By knowing much more about what people do, how they behave, how they respond to events or to policies, it becomes possible to generate predictions and forecasts about best possible courses of action, and then to intervene to either pre-empt how people behave or prompt them to behave in a certain way. Evelyn Ruppert et al (2017) have termed this 'data politics', noting that power over data no longer only belongs to bureaucracies of state, but to a constellation of new actors in different sectoral positions—companies, think tanks, consultancies, international organizations, data labs and so on. As such, earlier forms of state power performed through statistics are evolving in the era of big data as other non-state agencies take responsibility for data collection and analysis.

Something of an arms race is underway by those organizations that want to attain data power in education. Education businesses, venture capital firms and philanthropies are putting large financial, material and human resources into technologies of datafication, and are seeking both to make it commercially profitable and also attractive to policymakers as a source of intelligence into learning processes. Having 'ownership' over educational big data is potentially valuable as a way of building new technologies and gaining political traction. Dorothea Anagnostopoulos et al (2013) have written about the 'informatic power' possessed by the organizations and technologies involved in processing test-based data. But some of that power is now being assumed by those actors, organizations and analytics technologies that process digital learning data and turn it into actionable intelligence and adaptive, personalized prescriptions for pedagogic intervention. Organizations such as the Chan-Zuckerberg Initiative and the Gates Foundation are amongst the most powerful in contemporary data-driven education, partly through creating strategic partnerships with key policy influencers and partly by creating generous funding programs. The Chan-Zuckerberg Initiative, for example, recruited the former Deputy Secretary of the US Department of Education to head its education division, and has quickly become a multi-billion dollar funder of new education programs that align with its mission.

Governmental authority does not, of course, automatically confer trust on the findings generated from data. The UK's Department for Education has been repeatedly reprimanded by official statistics authorities for misusing its own datasets, notably for putting political spin and 'messaging' well ahead of the statistical evidence produced by its own in-house experts (Shah 2018). As Davies (2017) has noted, although statistical and analytical expertise is commonly regarded as a source of governmental authority, the power to narrate the data to make it into meaningful messages for the public and the media can often lead to misleading uses of numbers. Moreover, for Davies (2017), expert power in the era of datafication does not reside purely in human actors and social organizations; instead, data power is distributed between machines that can detect patterns in massive datasets and humans with the talent to narrate those patterns and produce conviction in others that the data are meaningful and truthful. The objectivity of the data and the algorithms invest authority in their human mediators—although the Department for Education example signals how misleading messaging can also undermine trust in the numbers.

As an example of how human subjective judgment is displaced by algorithmic objectivity, the Behavioural Insights Team recently ran an experimental program using 'school evaluating algorithms' as part of school inspection processes normally undertaken by human inspectors from Ofsted, the agency for standards in UK schools (Sanders et al 2017). Both the BIT and Ofsted are semi-governmental agencies. The aim of the trial, however, was not to replace the embodied human inspector, but to augment the inspection process by allowing the algorithm to identify problems in school data that could then help shape the approach taken by the human inspector. Teachers have for years voiced concern about the judgment of school inspectors and the effects on how they think about pedagogy and curriculum; now they must also adjust to the measures of the school-evaluating algorithm. As these examples indicate, data power in education is being distributed across commercial, philanthropic and non-governmental organizations that possess the objective analytical technologies and the human resources required to undertake complex data analyses and communicate their meanings.

Cultural data
Datafication is a cultural phenomenon and a concept that has attained a privileged position in the view of the public, businesses, governments and the media (Beer 2016). Increasingly, it seems, data and algorithms are invested with promises of objectivity and impartiality, at a time when human experts are not necessarily to be trusted because they are too clouded by subjective opinion, bias and partiality. Across the ed-tech sector, the apparent objectivity of data has been culturally adopted and accepted, based on the assumption that teachers are too subjectively biased and are unable to adequately keep track of all students' progress. This speaks to a cultural narrative framing datafication in terms of mechanical objectivity, certainty, impartiality, and framing human subjective judgment in terms of standpoint bias (Williamson & Piattoeva 2018).

The cultural acceptance or otherwise of datafication is of course context-specific. In some European countries such as Germany the cultural narrative of datafication and algorithms is more contested, and legally and politically influenced, due to greater cultural sensitivity around data privacy and protection. The 2018 General Data Protection Regulation (GDPR) in Europe is likely to exert significant effects on the ways that data are used in the EU, and open up distinctive differences in processes of datafication with other geopolitical spaces. It will influence how datafication in general and datafication of education in particular becomes culturally embedded (or not) in different geographical, political and social locations. The structure and organization of national education systems is also likely to impact on the ways data are collected and used. Artificial intelligence and facial recognition technologies for use in education are at an advanced state of development and implementation in China, for instance, reflecting both national economic imperatives around ed-tech innovation and low political priorities around personal privacy alike.

The rise of an ed-tech industry in China challenges the Anglophone dominance of the US and UK in datafication to date. The lobbying group EdtechUK, for example, has worked hard to make the UK into a world-leading centre of education technology development, while Silicon Valley already has an established innovation culture and infrastructure of edu-hackathons, accelerator programs and funding streams to incentivize ed-tech businesses and startups. The Anglophone culture of technology innovation, however, produces particular approaches to education, learning and pedagogy that may not be culturally appropriate in other contexts and settings, especially in the Global South (Slade & Prinsloo 2013). Recently, a new report on 'Learning Analytics for the Global South' considered 'how the collection, analysis, and use of data about learners and their contexts have the potential to broaden access to quality education and improve the efficiency of educational processes and systems in developing countries around the world' (Lim & Tinio 2018). The papers in the report suggest that datafication of education is becoming increasingly culturally sensitive.

At the same time, however, global organizations such as UNESCO, OECD and Global Partnerships for Education are seeking to produce new statistical measures to capture basic educational data from developing countries, ultimately casting a grid of standardized metrics over international development contexts and glossing over cultural diversity in the efforts to globalize educational measurement and comparison (Verger, Novelli & Altinyelken 2018) As such, the enactment and effects of datafication on educational policy and pedagogy need to be understood as embedded in cultural context, whilst also acknowledging that many approaches to datafication may be decontextualized, universalistic, and privileging of technocratic Westernized ideals. Ambitious Chinese expansion, though, suggests new culturally-located expressions of datafication are emerging to challenge Anglophone dominance.

Data legalities & ethics

Finally, there are legal, ethical and regulatory mechanisms shaping datafication. Europe is much more privacy-focused than the US or China, for example, as the EU GDPR shows. How datafication plays out—what datafication *is*—is itself shaped by law, ethics and politics. Even without GDPR in the US, specific federal acts such as COPPA and FERPA exist to protect children's privacy, and organizations like the Internet Keep Safe Coalition enforce them (Zeide 2016). Other organizations such as the Future of Privacy Forum exist to produce 'policy guidance and scholarship about finding the balance between protecting student privacy and allowing for the important use of data and technology in education'.

Education policy also shapes the legal environment in which datafication can occur. The US 2015 Every Student Succeeds Act (ESSA) has made it possible for states and schools to apply for additional funding for personalized learning technologies. The new federal act performs the double task of stimulating market growth in adaptive personalized learning software and incentivizing schools to invest in such technologies in the absence (or at least shortage) of public funding for state schooling. As such, ESSA makes datafication of public education possible at huge scale, and even financially incentivizes schools to invest in data platforms. In higher education in the UK, meanwhile, governmental changes in the scope and scale of data collection about students mean that new markets are opening for commercial providers of 'data services solutions' (Komljenovic 2018).

Of course, the ethical issues of datafication of education are considerable. School cybersecurity incidents are already a routine occurrence in the US. Major ed-tech companies including EdModo and Chegg have experienced huge data breaches through hacking attacks. Education companies including Blackboard and Pearson have been the subject of fierce backlash after appearing to use student data for secondary purposes without notice or consent. As GDPR compliance became a legal requirement in 2018, US education companies with EU users were forced to rework their privacy policies and data sharing agreements with third party service providers. Nonetheless, considerable unresolved concerns remain about the adequacy of contemporary student privacy and data protection policies and frameworks in relation to the rise of educational datafication (Ziede 2016).

Conclusion

This chapter has offered a series of defining aspects of datafication and raised issues and challenges regarding education, pedagogy, and curriculum design. Its purpose was to open up educational data to debate and further scrutiny, as the uses of data in education are likely to intensify and escalate over coming years. None of this is to discount the possibility of meaningful uses of data to inform pedagogy and curriculum. Nor is it to treat datafication as inevitable and deterministic. Research fields such as learning analytics and education data science are continuing to develop and refine their approaches, including close attention to a range of ethical issues. The critical social scientific issues outlined in this chapter, however, point to the enduring need for alternative perspectives that are attentive to the history, technicalities, epistemology, social consequences, cultural contingency, and politics of datafication in education. Without such perspectives, educational understanding of datafication would remain limited to its practical role in teaching and neglect its wider social, cultural, economic, and political significance.

References

AI Now. 2018. *AI Now Report 2018*. New York: AI Now Institute.

Ambrose, M. 2015. Lessons from the avalanche of numbers: big data in historical perspective. *I/S: A Journal of Law and Policy for the Information Society* 11, no. 2: 201-277.

Anagnostopoulos, D., Rutledge, S.A. & Jacobsen, R. (eds). 2013. *The Infrastructure of Accountability: Data use and the transformation of American education*. Cambridge, MA: Harvard Education Press.

Beer, D. 2016. *Metric Power*. London: Palgrave Macmillan

Boninger, F. & Molnar, A. 2016. *Learning to be Watched: Surveillance Culture at School.* Boulder, CO: National Education Policy Center.

Bowker, G.C. 2013. Data flakes: an afterword to *'Raw Data' is an Oxymoron*. In Gitelman, L. (ed.) *'Raw Data' is an Oxymoron*: 167-172. London: MIT Press

boyd, d. & Crawford, K. 2013. Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15, no. 5: 662-679.

Bradbury, A. & Roberts-Holmes, G. 2018. *The Datafication of Primary and Early Years Education: Playing with numbers*. London: Routledge.

Bulger, M. 2016. Personalized learning: the conversations we're not having. *Data and Society*, 22 July 2016: https://www.datasociety.net/pubs/ecl/PersonalizedLearning_primer_2016.pdf

Cope, B. & Kalantzis, M. 2015. Interpreting Evidence-of-Learning: Educational research in the era of big data. *Open Review of Educational Research* 2, no. 1: 218-239.

Davies, W. 2017. Elite Power under Advanced Neoliberalism. *Theory, Culture & Society*: http://dx.doi.org/10.1177/0263276417715072

DiCerbo. K. E. & Behrens, J. T. 2014. *Impacts of the Digital Ocean*. Austin, TX: Pearson.

Finn, M. 2016. Atmospheres of progress in a data-based school. *Cultural Geographies* 23, no. 1: 29–49.

Foucault, M. 1991. *Discipline and Punish: The Birth of the Prison*. Trans. A. Sheridan. Harmondsworth: Penguin.

Gehl, R. 2015. Sharing, knowledge management and big data: A partial genealogy of the data scientist. *European Journal of Cultural Studies* 18, no. 4-5: 413-428.

Gillespie, T. 2014. The Relevance of Algorithms. In Gillespie, T., Boczkowski, P. J., & Foot, K. A. (eds). *Media technologies: Essays on communication, materiality, and society:* 167-193. London: MIT Press.

Gitelman, L. & Jackson, V. 2013. Introduction. In L. Gitelman (ed.) *'Raw Data' is an Oxymoron*: 1-14. London: MIT Press.

Gorur, R., Sellar, S. & Steiner-Khamsi, G. (eds). 2019. *Comparative Methodology in the Era of Big Data and Global Networks*. London: Routledge.

Gray, J. 2016. Datafication and democracy. *Juncture*, 21 December: https://www.ippr.org/juncture/datafication-and-democracy

Hill, P. & Barber, M. 2014. *Preparing for a renaissance in assessment*. London: Pearson.

Housley, W. 2015. The Emerging Contours of Data Science. *Discover Society* 23: http://discoversociety.org/2015/08/03/focus-the-emerging-contours-of-data-science/

Kitchin, R. 2014a. *The Data Revolution: Big data, open data, data infrastructures and their consequences*. London: Sage.

Kitchin, R. 2014b. Big Data, new epistemologies and paradigm shifts. *Big Data & Society* 1, no. 1: http://dx.doi.org/10.1177/2053951714528481

Komljenovic, J. 2018. Making higher education markets: trust-building strategies of private companies to enter the public sector. *Higher Education*: https://doi.org/10.1007/s10734-018-0330-6

Lang, C., Siemens, G., Wise, A. & Gasevic, D, (Eds) (2017). *Handbook of Learning Analytics*, 1st edn. Society of Learning Analytics Research.

Lim, C. P., & Tinio, V. L. (Eds.). (2018). Learning analytics for the global south. Quezon City, Philippines: Foundation for Information Technology Education and Development.

Lupton, D. 2015. *Digital Sociology*. London: Routledge.

Lyotard, J.-F. 1984. The Postmodern Condition: A report on knowledge. Trans. G. Bennington & B. Massumi. Manchester: Manchester University Press.

Manolev, J., Sullivan, A. & Slee, R. 2018. The datafication of discipline: ClassDojo, surveillance and a performative classroom culture. *Learning, Media & Technology*: https://doi.org/10.1080/17439884.2018.1558237

Mulgan, G. 2016. Government as collective intelligence. *Oxford Government Review* 1, August 2016: 44-46.

Piety, P.J., Hickey, D.T. & Bishop, M.J. 2014. Educational data sciences—framing emergent practices for analytics of learning, organizations and systems. *LAK '14,* March 24 - 28 2014, Indianapolis.

Slade, S. & Prinsloo, P. 2013. Learning Analytics: Ethical Issues and Dilemmas. *American Behavioural Scientist* 57, no. 10: 1510-1529

Roberts-Mahoney, H. Means A.J. & Garrison, M.J. 2016. Netflixing human capital development: personalized learning technology and the corporatization of K-12 education. *Journal of Education Policy*: http://dx.doi.org/10.1080/02680939.2015.1132774

Ruppert, E. 2012. The Governmental Topologies of Database Devices. *Theory, Culture & Society* 29, no. 4-5: 116-136.

Ruppert, E., Isin, E. & Bigo, D. 2017. Data politics. *Big Data & Society*, July-December: 1-7.

Sellar, S., Thompson, G. & Rutkowski, D. 2017. *The Global Education Race: Taking the measure of PISA and international testing*. Brush Education.

Selwyn, N. 2015. Data entry: towards the critical study of digital data and education. *Learning, Media & Technology* 40, no. 1: 64-82.

Shah, H. 2018. The DfE's repeated misuse of statistics is embarrassing. Schools Week, 8 October: https://schoolsweek.co.uk/the-dfes-repeated-misuse-of-statistics-is-embarrassing/

Silberzahn, R., Uhlmann, E.L., Martin, D.P. et al. 2018. Many Analysts, One Data Set: Making Transparent How Variations in Analytic Choices Affect Results. *Advances in Methods and Practices in Psychological* Science 1, no. 3: 337–356.

van Dijck, J. 2014. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society* 12, no. 2: 197-208.

Verger, A., Novelli, M., & Altinyelken, H.K. (eds). 2018. *Global Education Policy and International Deevelopment: New agendas, issues and policies*, 2nd ed. London: Bloomsbury.

Williamson, B. 2017. *Big Data in Education: The digital future of learning, policy and practice*. London: Sage.

Williamson, B. 2018. The hidden architecture of higher education: building a big data infrastructure for the 'smarter university'. *International Journal of Educational Technology in Higher Education* 15(12): https://doi.org/10.1186/s41239-018-0094-1

Williamson, B. & Piattoeva, N. 2018. Objectivity as standardization in data-scientific education policy, technology and governance. *Learning, Media and Technology*: https://doi.org/10.1080/17439884.2018.1556215

Zeide, E. 2016. Student Privacy Principles for the Age of Big Data: Moving Beyond FERPA and FIPPS. *Drexel Law Review* 8: 339