Smarter learning software:
Education and the big data imaginary

Ben Williamson, University of Stirling

Paper prepared for *Big Data—Social Data*, University of Warwick, 10 December 2015

Abstract Big data and smarter learning software systems are beginning to impact on education, particularly within the schools sector. This paper traces the emergence of a 'big data imaginary,' a vision of a desirable future of education that its advocates believe is attainable through the application of big data technologies and practices. Firstly, it identifies a 'first wave of big data' in nineteenth-century education exhibitions and its continuities with the visualization of large-scale educational data today. Secondly, it details the emergence of 'educational data science' as an exemplar of how 'second wave big data' has entered the imagination of many actors within education. Thirdly, it then demonstrates how education is being reimagined in relation to 'smart cities' that depend on big data for their functioning, before fourthly detailing the recent appearance of 'startup schools' that are being established by Silicon Valley entrepreneurs to run as testbeds of smarter learning software systems. A concluding section discusses how the future of education may be governed by the production and circulation of the 'data and algorithms of the powerful.'

Keywords big data, imaginaries, schools, software

Education can be thought of as a social institution that is centred on 'smartness.' Etymologically, 'smart' means to be educated, clever, intelligent, knowledgeable, capable, adept, and quick at learning. But education is also now increasingly augmented by 'software smarts'—technical devices with the appearance of a degree of intelligence, programmed to be capable of some independent action, and the capacity to react or respond to differing requirements, varying situations, or past events. In particular, with the emergence of 'big data,' education as a social institution is beginning to encounter smart software that can itself learn from the data it processes and analyses—through processes such as machine learning, whereby algorithms are 'trained' on data to perform the desired calculations of their producers (Mackenzie 2015)—and then respond to those data independently and intelligently. 'Smarter learning software' is the term I use to describe such emerging technologies and devices, and the human and technical practices that enact them.

In this paper I discuss some of the ways in which the rise of big data and smarter learning software systems are impacting on education, with a particular focus on emerging developments in the schools sector. From the outset, I acknowledge the contested definitional nature of big data, as captured in the question 'What makes big data, big data?' by Kitchin and McArdle (2015), who note that many systems described as big data often fail to meet the definitional criteria offered by the literature (e.g. huge in volume, highly varied, collected in real-time, able to be extended in scale and related to other datasets, exhaustive rather than 'sampled,' and fine-grained in resolution). Rather, big data has become 'loose in its ontological framing and definition,' often 'treated like an amorphous entity that lacks conceptual clarity,' and 'while there has been some rudimentary work to identify the "genus" of big data … there has been no attempt to separate out its various "species" and their defining attributes' (Kitchin & McArdle 2015: 4-5). In other words, there may be many kinds of big data with different characteristics and nature—as is the case in education itself.

Instead of seeking to define what makes an educational big data system big data, however, I trace the emergence of what I term a 'big data imaginary' for education. This term is adapted from Jasanoff (2015) who describes a 'sociotechnical imaginary' as a collectively held, institutionally stabilized, and publicly performed vision of a desirable future that is animated by shared understandings of forms of social life and social order and made attainable through the design of technological projects. Such futures are produced by particular social groups within specific social contexts, and they are also projected through the design of particular kinds of technologies to express a view of particular futures in which those kinds of technologies are imagined to be integral, embedded parts. As Jasanoff (2015: 5-6) elaborates:

> It often falls to … institutions of power to elevate some imagined futures above others, according them a dominant position…. Imaginaries, moreover, encode not only visions of what is attainable through science and technology, but also of how life ought, or ought not, to be lived.

Thus the dreamscapes of the future that are dreamt up in technical R&D departments, science labs, entrepreneurs' studios, and so on, sometimes, through collective efforts, become stable and shared objectives that are used in the design and production of actual technologies and scientific innovations—developments that then incrementally materialize the desired future. In slightly different terms, the imagining of a 'digital future' projects a kind of 'mythology' (a set of ideas and ideals) that animates, motivates and drives forwards technical development, but is always much more contested and 'messy' than it is imagined to be (Dourish & Bell 2011). For Housley (2015) a latent goal of the 'Big Data Imaginary' (his capitalization) is to configure 'social organization and relations according to mathematical principles' derived from 'the empirical focus of data science.' The big data imaginary of an institution of power such as data science, then, motivates excursions into the empirical, and seeks to make social worlds amenable to its underlying goals and aspirations.

As such, it is important to consider the 'data practices' that generate digital data, and to acknowledge that the ways these data are interpreted and made meaningful are also *generative* of particular effects and social implications, since data and the algorithms that process them are consequential to 'what is known,' and can influence decision-making and other activities (Ruppert *et al.* 2015). Throughout this paper I treat data practices as social and technical instantiations of particular future visions—as practices that operationalize imaginaries in the present, often in ways that are more messy in practice than anticipated or desired. By mobilizing the idea of a big data imaginary along with the emphasis on data practices, what I am seeking to register is the formation and circulation of a 'desirable' educational future associated with (and made attainable through) data practices, technologies, discourses, knowledges, institutions and cultures. The result of such an imaginary is the proposed production of smarter learning software systems that might transform educational practices, processes, policies and institutions, although the materialization of such an imaginary of education in practice is likely to be more contested and messy than its advocates might imagine.

The paper is in five further parts. Firstly, I identify a 'first wave of big data' in nineteenth-century educational exhibitions, and its scaling-up through more recent large-scale databanks. The second part details the emergence of 'educational data science' as an exemplar of how a 'second wave of big data' has entered the imagination of many actors within education. Thirdly, I then demonstrate how education is being reimagined in relation to 'smart cities' that depend on big data for their functioning, before fourthly detailing the recent appearance of 'startup schools' that are being established by Silicon Valley entrepreneurs to run as testbeds of smarter learning software systems. The fifth part then concludes with a brief discussion considering how the variety of data practices now being undertaken may be making the big data imaginary a practicable reality for educators and learners to inhabit. While sociological studies of education have long been concerned with how 'knowledge of the powerful' is reproduced through schooling, I consider whether the introduction of smarter learning software means that the schools of the future may be governed by the 'data and algorithms of the powerful.'

Exhibiting educational data

The collection, calculation and communication of data in education has a long history. Numerical data were already being mobilized in both Europe and the United States as part of the spectacle and display of the great exhibitions, world's fairs and scientific congresses of the nineteenth century (Lawn 2013). At these events, the visualization of large quantities of data through graphical displays such as topographical maps, tables of illustrative statistics, and exemplars of student's work could be used to represent the relations between education and society directly to the viewer.

The nineteenth-century datafication of education was part of the 'avalanche of numbers' that Hacking (1990) associated with the rise of statistics and other novel knowledge production and sorting processes of the mid-1800s, such as those of census offices, libraries, and museums, and especially those statistical practices associated with counting rates of sickness, disease, poverty, crime, and so on. Ambrose (2014: 17) describes the avalanche of numbers as 'first wave big data,' as 'people reached for quantification when chaos from a massive shift in the sociotechnical world ensued' from the industrial revolution. This had significant implications for social organization, control and governance. Robertson and Travaglia (2015) have noted that the nineteenth-century 'collection of social data had a purpose—understanding and controlling the population in a time of significant social change'—which brought about changed understandings of individuals, groups and populations, and influenced how they might be acted upon through new social institutions.

As Sobe (2013) has documented, the original world's fairs expositions played a major role in formulating many modern social and institutional structures, and from the nineteenth century education—as one such institution—was already being exhibited, made visible and comparable through the display of data as 'clear illustrations' to external audiences at these events. These images of the data, the numerical information they represented, and the statistical techniques used to generate them, were powerful explanatory and persuasive devices which made it difficult to conceive of education in any other terms—and could be used as the basis for political intervention. Sobe concludes that the 'exhibitionary practices' of the nineteenth century are now being continued in the 'scopic' techniques mobilized in the spectacular display of educational data today, such as the printed reports, electronic files and spreadsheets that are distributed from government education departments and non-governmental agencies such as the OECD (Organization of Economic Cooperation and Development). The OECD has itself become a major producer of large-scale and longitudinal educational data from its standardized global tests for both children and adults (Lawn 2013). As a social institution, education has long reached for quantification and the exhibition of its numbers as an explanatory and rhetorical source.

The continuing statistical and exhibitionary practices associated with the collection, calculation and communication of educational data today can be seen clearly in two examples. The National Pupil Database (https://www.gov.uk/government/collections/national-pupil-database) was established by the UK government in 2002. It features extensive datasets on the educational progress of children and young people from the early years through to higher education, and includes data on seven million pupils matched over twelve years—including pupil identifiers, home addresses, test and exam results, prior attainment and progression, pupil characteristics such as gender, ethnicity, first language, free school meals, special educational needs, absences and exclusions—and is also connected to databases holding similar information from further and higher education to produce linked data. The data are drawn from regular school censuses (usually conducted three times a year) as well as from Local Authorities and awarding bodies, and are processed by the Department for Education's Education Data Division and matched and stored in the NPD. The NPD is presented in Excel spreadsheet files as thousands upon thousands of rows of numbers that can be searched and analyzed in myriad ways by government departments (as well as by authorized third parties), and used to generate complex and sophisticated graphical displays such as charts, tables, plots, graphs and so on. It is, in other words, a statistical device for exhibiting large-scale educational data in order to inform governmental policy and analysis.

The NPD was launched by Michael Barber while a champion of evidence-based policymaking in the Prime Minister's Delivery Unit. Ten years later, in his new role as chief education adviser to the commercial education publisher Pearson, Barber launched the Learning Curve (http://thelearningcurve.pearson.com/), a much more dynamic, interactive and visual exhibition of educational data at a global level. The Learning Curve combines over 60 educational datasets from around the world into one place in order to 'enable researchers and policymakers to correlate education outcomes with wider social and economic outcomes.' The Learning Curve includes national performance data (sourced from, for example, the National Pupil Database) along with the OECD's PISA data and other sources such as UNESCO, in order to produce a 'Global Index' of nations that is ranked in terms of 'educational attainment' and 'cognitive skills' by country. The Learning Curve is highly relational, enabling the conjoining of multiple datasets, as well as scalable in that it can expand rapidly—new data is added frequently as new datasets from its various sources become available. It is especially notable for its data visualization tools. It features a suite of dynamic, interactive and user-friendly mapping and time series tools that allow countries to be compared and evaluated both spatially and temporally. Countries' educational performance in terms of educational attainment and cognitive skills are represented on the site as semantically resonant 'heat maps.' It also permits the user to generate 'country profiles' that visually compare multiple 'education input indicators' (such as public educational expenditure, pupil:teacher ratio, educational 'life expectancy') with 'education output indicators' (test scores, graduation rates, labour market productivity), as well as 'socio-economic indicators' (such as GDP and crime statistics). The Learning Curve is a powerful technique of political visualization for envisioning and exhibiting the educational landscape, enabling numbers to be presented and re-presented to function for a variety of purposes, users and audiences. Hogan, Sellar and Lingard (2015: 7) describe it as 'attractive to policymakers because it simplifies complex problems in education' and also 'potentially opens a new commercial space for Pearson to generate profit through the provision of policy solutions.'

The NPD and the Learning Curve are indicative of how the avalanche of numbers and the exhibition of data in the nineteenth century are being continued in the current enthusiasm for quantification and visualization in education, and particularly of how the situating of such data within software systems permits users to engage in their own interactive exhibitionary practices. Users of the NPD and (especially) the Learning Curve are solicited to become data practitioners and data co-producers whose own analyses and interpretations are structured by the interactive software interface (Williamson 2016). Their datasets have powerful explanatory force, and construct virtual worlds of educational data (Lawn 2013) that can be used rhetorically to produce conviction in others that the only way to see and know education as a social institution is through its numbers and its graphical display (Williamson 2015a).

Educational data science

If 'first wave big data' in education could be characterized by the great expos, and continued in contemporary software devices that make educational numbers even more visible, interactive and reproducible, then 'second wave big data' in education is captured by the emergence of a new field of expertise known as 'educational data science.' Piety, Hickey and Bishop (2014) have defined educational data science in terms of its combination of 'Academic/Institutional Analytics,' 'Learning Analytics/Educational Data Mining,' 'Learner Analytics/Personalization,' and 'Systemic Instructional Improvement,' all of which are 'significantly related to digital technology and its ability to collect, share, and represent vast amounts of information with relative ease.' These four areas of research and development, they argue, have begun to coalesce around shared questions, problems and assumptions over the last decade to form a field that 'has begun to receive combined attention from both federal policymakers and foundation funders and is often seen as the community dealing with "Big Data" in education.'

These authors term educational data science a 'sociotechnical movement' with shared interests that cut across the boundaries of their original communities. By sociotechnical movement what they mean is that 'the enabling conditions and key technologies emerge across sectors giving rise to multiple sets of innovations that may at times seem disconnected, but are often related and interdependent.' They also point out that a sociotechnical movement can gain traction when society's 'expectations are such that the innovations come at a time when there is other general interest in the kinds of changes that the innovations make possible.' Thus they note how there has, in recent years, been both increasing capability to produce data and a greater public appetite for the use of data across many areas of education. They also highlight how new forms of evidence—log files, conversational records, peer assessments, online search and navigation behaviour, and others—are raising big questions and disrupting traditional ways of working in educational research, 'acting in a way similar to *disruptive innovations* that alter cultural, historical practices and activity systems.'

The notion that educational data science, as the community most involved in big data in education, may be radically disrupting the social institution and practices of education itself is captured in an e-book by Mayer-Schönberger and Cukier (2014) entitled *Learning with Big Data: The Future of Education.* The authors suggest that big data will 'reshape learning' through 'datafying the learning process' in three significant ways: through real-time feedback; individualization and personalization of the educational experience; and probabilistic predictions to optimize what students learn. These changes are being brought about, they argue, through a combination of:

- Online courses that enable the constant logging and tracking of learners through their clickstream data
- E-textbooks that can 'learn' from how they are used and 'talk back' to the teacher
- Adaptive learning systems that enable materials to be tailored to each student's individual needs through automated real-time analysis
- The generation of personalized 'playlists' determined by an algorithm in order to 'optimize how people learn'
- New forms of data analytics that are able to harvest data from students' actions, learn from them, and generate predictions of individual students' probable future performances
- Automated personal tutoring software that monitors students and gives constant real-time support and shapes the pedagogic experience

The publication provides a seamless image of school as a 'data platform,' the 'cornerstone of a big-data ecosystem' in which 'educational materials will be algorithmically customized' and 'constantly improved,' all accomplished via educational 'algorithmists' and other data intermediaries, professionals trained in the science of big data who can advise on data policies and methods. *Learning with Big Data* is evidence of how education in the near future is being imagined as organized through the educational data science practices of data collection, analysis and presentation.

What do such developments mean in terms of professional expertise? Buckingham Shum and colleagues (2013) have identified that 'while the learning analytics and educational data mining research communities are tackling the question of what data can tell us about learners, relatively little attention has been paid, to date, to the specific mindset, skillset and career trajectory of the people who wield these tools.' They term educational data scientists a 'scarce breed,' and note in particular that they would need to be experts in both learning analytics and educational data mining, as well as in a host of related techniques. In defining the field Piety, Behrens and Pea (2014) traced its disciplinary origins to computer science techniques of computational statistics, data mining, machine learning, natural language processing and human-computer interaction.

Further taking up the challenge of defining the mindset and skillset of educational data scientists, Pea (2014) has proposed a new 'specialized' field combining the sciences of digital data and the science of learning, and the construction of a 'big data infrastructure' for learning consisting of data science and computer science techniques that could be harnessed to the challenge of analysing large volumes of educational and learning data. Specifically, his report identifies 'several competencies for education data science,' including:

- Computational and statistical tools and inquiry methods, including traditional statistics skills … as well as newer techniques like machine learning, network analysis, natural language processing, and agent-based modeling
- General educational, cognitive science, and sociocultural principles in the sciences of learning…
- Principles of human–computer interaction, user experience design, and design-based research
- An appreciation for the ethical and social concerns and questions around big data, for both formal educational settings and non-school learning environments

Likewise, DiCerbo and Behrens (2014), of the commercial educational publisher and software vendor Pearson, have avidly advocated a new datafied science of learning, arguing that as 'teaching and learning becomes digital, data will be available not just from once-a-year tests, but also from the wide-ranging daily activities of individual students … in real time. … [W]e need further research that brings together learning science and data science to create the new knowledge, processes, and systems this vision requires.'

These presentations and reports clearly demand a lot of expertise from educational data scientists. Pea's (2014) report calls for much more support from governments for this sector, and details the need for new undergraduate and graduate courses to support its development. Pearson, for its part, has established a Center for Digital Data, Analytics and Adaptive Learning (http://researchnetwork.pearson.com/digital-data-analytics-and-adaptive-learning) where it is practising educational data science in-house. Its director John Behrens is an expert in measurement and statistics, whose research focuses on how 'the billions of bits of digital data generated by students' interactions with online lessons as well as everyday digital activities can be combined and reported to personalize learning,' while other staff in the center are described as 'research scientists' with expertise in data mining, computer science, algorithm design, intelligent systems, human-computer interaction, data analytics tools and methods, and interactive data visualization.

Educational data science is not, then, simply a technical field of expertise in statistical forms of analysis, but is deeply rooted in 'learning science,' a field itself largely defined in terms of concepts and methods from the psychological and cognitive sciences. The expertise of an educational data scientist is a hybrid of computer science (CompSci) and the psychological sciences (psy-science). I have elsewhere referred to the combination of CompSci and the psy-sciences as a 'CompPsy' interdisciplinary hybrid and argued that the juxtaposition of computational methods of analysis with psychological concepts is giving rise to new theories and understandings of learning that appear to challenge the accounts offered by educational researchers that are based on empirical fieldwork, ethnographies and other situated methodologies (Williamson 2016). Instead, CompPsy practices take place in a microlaboratory inside a computer system, using data analysis techniques and practices to detect patterns in the millions of digital traces left when users undertake a task or activity. The digital microlaboratories of educational data science are both small enough to be written in computer code, but also massively distributed to aggregate individuals' data into big population datasets that can be analysed for patterns at huge scale.

This raises important questions about the actual subjects of the research conducted by educational data scientists, and the potential insights that can be drawn from analyzing their data. DiCerbo and Behrens (2014) of Pearson's Center for Digital Data, Analytics and Adaptive Learning argue that as learners interact with systems and with other people, 'software records' every aspect of their activity, with the result that:

> these developments have the potential to inform us about patterns and trajectories for individual learners, groups of learners, and schools. They may also tell us more about the processes and progressions of development in ways that can be generalised outside of school.

The promise of the educational data science methods being pioneered and practised by Pearson is therefore not simply of better tracking of learners but the generation of new generalizable theories and models of cognitive development and learner progression. Likewise, Pea (2014) has highlighted 'a pre-eminent objective' in educational data science of 'creating a model of the learner'—a model that can then be used as the basis to generate predictions about probable future progress, and to inform future pedagogic intervention.

In another Pearson paper on the methodological challenges of analyzing educational big data, Behrens (2013) claims insights extracted from the generation of huge quantities of educational data will challenge current theoretical frameworks in education research, as 'new forms of data and experience will create a theory gap between the dramatic increase in data-based results and the theory base to integrate them.' The CompPsy laboratories of educational data science focus on models and patterns derived from digital trace data, and mobilize those patterns and models in the construction of new theories of learning itself. Such practices and methods relocate the subjects of educational research from situated settings and psychological labs to the digital laboratory inside the computer, and in doing so transform those subjects from embodied individuals into numerical patterns, data models, and visualized artefacts. Companies like Pearson may well then be able to use those data as insights in the production of new e-learning software products that can be marketed to schools and colleges. This would amount to the encoding of new theories of learning into marketable software products that would shape the subsequent tasks—and thus the learning processes—of potentially large numbers of students.

In sum, educational data science is an emerging multidisciplinary field of technical and methodological expertise, one that has developed as a movement across academic and commercial settings through a family tree of influences and practices, and that has significant potential to transform aspects of educational research and theory in future years—especially if funding and governmental backing accumulate to support its ambitions. For example, as the world's largest educational publisher and e-learning provider, Pearson is inserting educational data science practices and methods into schools and colleges worldwide. Notably, Pearson has partnered with Knewton, a major learning analytics provider, to power its digital content:

> The Knewton Adaptive Learning Platform™ uses proprietary algorithms to deliver a personalized learning path for each student…. 'Knewton adaptive learning platform, as powerful as it is, would just be lines of code without Pearson,' said Jose Ferreira, founder and CEO of Knewton. 'You'll soon see Pearson products that diagnose each student's proficiency at every concept, and precisely deliver the needed content in the optimal learning style for each. These products will use the combined data power of millions of students to provide uniquely personalized learning.' (http://www.knewton.com/press-releases/pearson-partnership/)

Knewton's Adaptive Learning Platform has been detailed on the company website as enabling 'real-time assessment of the individual,' and enacts techniques such as 'algorithmic assessment norming at scale,' sophisticated database architecture and tagging, complex taxonomic systems, 'groundbreaking machine learning algorithms,' 'inferred student data,' and prescriptive analytics 'recommender systems.'

As Knewton's platform and Pearson's products scale out globally, each constructed to perform an automated diagnostics on every individual learner and then algorithmically personalize their learning pathways, we can see how a virtual microlaboratory of educational data analysis may be installed directly in the smarter learning software with which students interact. As digitized learning environments increasingly become the norm (and Pearson and Knewton aspire to make this a reality), young people will be learning in the digital microlaboratory itself, the constant subjects of diagnostic learning analytics and adaptive learning platforms. This comes with risks. As the 'robotic algorithms' of learning analytics platforms are able to access spreadsheets of learner data, calculate odds and make probabilistic predictions, and automate decisions about pedagogical intervention in a few milliseconds, 'the risk [is] that our predictions may, in the guise of tailoring education to individual learning, actually narrow a person's educational opportunities to those predetermined by some algorithm' (Mayer-Schönberger & Cukier 2014). As Hope (2014) notes, the insights constructed from an individual's data set effectively serve to either permit or inhibit educational opportunities.

The digital microlaboratories of educational data science are new sources of knowledge production with direct access to research subjects at very large scale. Educational data science is embryonic of second wave big data in education, and exemplifies how a shared imaginary of education augmented and optimized via big data is already being operationalized and materialized in the present. The spread of such an imaginary of the future of education is especially evident in relation to current reimaginings of education for 'smart cities.'

Educating smart cities

In the fields of architecture, computational urbanism, and urban science, increasing attention has turned in recent years to the emergence of 'smart cities,' urban environments augmented with 'big data,' 'sensor networks,' 'ubiquitous computing,' 'coded infrastructures' and other computationally programmable processes and software-supported data practices scripted in code (e.g. Batty 2013; Shepard 2011; Townsend 2013; Verebes 2014), As Batty and coauthors (2012: 482) define this 'visionary approach':

> The convergence of information and communication technologies is producing urban environments that are quite different from anything that we have experienced hitherto. Cities are becoming smart not only in terms of the way we can automate routine functions serving individual persons, buildings, traffic systems but in ways that enable us to monitor, understand, analyse and plan the city to improve the efficiency, equity and quality of life for its citizens in real time.

Along these lines, commercial computing firms have launched projects promoting their products for smart cities, including IBM, Cisco, Intel, Siemens and Microsoft, many linked to huge urban projects building new smart cities from the ground up, such as Songdo in South Korea, PlanIT Valley in Portugal, and Masdar in Abu Dhabi, while large funding grants have been awarded to research on digital urban infrastructures, many based at new research centres at universities, and political initiatives have made questions about the future of cities into a subject of governmental attention. A more critical appraisal of smart cities initiatives and discourses has been offered by geographers. Reviewing the smart cities literature, Gabrys (2014) characterizes them as urban spaces enabled by automated infrastructures; equipped with networked digital sensors and ubiquitous computing; that provide augmented experiences through mobile devices; and that mobilize the capture and analysis of 'big data' from urban processes in real time. As Kitchin (2014: 5) notes, these 'forms of instrumentation provide abundant, systematic, dynamic, well-defined, resolute, relatively cheap data about city activities and processes, enabling the possibility of real-time analytics and adaptive forms of management and governance.' The highly-instrumented and programmable 'real-time city' is one in which urban data analytics provide powerful means for making sense of and managing urban life, and for envisioning and predicting future scenarios for which pre-emptive plans can then be formulated and enacted. It is an urban environment structured and supported 'line by line, algorithm by algorithm, program by program,' 'by code using data as fuel' (Thrift 2014: 10), without which it would cease to function as planned.

Many of the major global companies associated with smart cities development have begun to address the question of how education might need to be reimagined for such software-smart urban spaces (for a more detailed account see Williamson 2015c). It is the potential of big data in particular that has captured the imagination of major commercial smart cities vendors such as Microsoft and IBM, whose visions of smart education systems depend on it. For example, the Microsoft Educated Cities program (http://www.microsoft.com/en-gb/citynext/education.aspx) provides a clear sense of how an imaginary of future schooling has been attached to big data developments, in that 'to be competitive, cities need to ensure that their citizens have access to twenty-first century productivity tools …, world-class apps and online services that make it easier to interact and collaborate.' Consequently, Microsoft advocates the use of 'analytics solutions' to gain insight from a range of educational data:

> Educators already have plenty of administrative and economic data—the challenge is gaining insights from it … [for] better planning and decision-making as well as improved tracking and evaluation. Microsoft and our partners create education analytics solutions that help students perform better and that can be adapted to meet individual needs. The analytics tools improve administration as well with a 360-degree view of performance and operations.

The IBM Smarter Education program (http://www.ibm.com/smarterplanet/us/en/education_technology/ideas/) is based on similar claims about the real-time availability of educational data:

> Schools and universities have always recorded and stored data as they tracked grades, attendance, test scores and demographics. With the increasing availability of technology in the instructional process, educational institutions now collect, in real time, data about what their students learn and how they progress … using big data and analytics.

Like the smart city itself, IBM's Smarter Education program is also premised on the assumption that data analytics will be one of the key technology 'trends' driving 'the future of learning':

> Analytics translates volumes of data into insights for policy makers, administrators and educators alike so they can identify which academic practices and programs work best and where investments should be directed. By turning masses of data into useful intelligence, educational institutions can create smarter schools for now and for the future.

In its vision of 'smarter schools,' IBM particularly emphasizes the use of 'academic analytics' to enable institutions to analyze data for insights into institutional effectiveness, and 'learning analytics' to facilitate the interpretation of students' actions, in ways that are wholly consistent with educational data science approaches.

An illustration of how such data analytics capacities may be embedded in education in the smart city is provided by the IBM Smarter Education vision for a 'smarter classroom' (http://www.ibm.com/smarterplanet/us/en/ibm_predictions_for_future/ideas/#Education). The IBM 'smarter classroom' is a 'classroom that will learn you' through 'cognitive-based learning systems' and both predictive and prescriptive analytics. Predictive tools, IBM claims, can answer the question: 'based on what's already happened, what's going to happen next?' Prescriptive analytics then answer: 'in light of what we believe is going to happen, what is the best response? These two dimensions of smarter analytics enable educational leaders to detect patterns that exist in masses of data, project potential outcomes and make intelligent decisions based on those projections.' The smarter classroom exemplifies how, according to the big data imaginary of IBM, schools will become able not only to provide real-time data on student activities, but also to make 'future-tense' predictions of their likely outcomes and to prescribe automated interventions that might nudge their individual and social behaviour and so pre-empt their futures.

To this end, IBM has established its own high school chain in the US, P-TECH (http://www.ptech.org/), the ambition of which is 'to build for schools what its operations center is for cities: a single system for collecting, aggregating and analyzing data from students and teachers alike, then writing algorithms to prescribe how to cope,' and a 'software "infrastructure layer" for schools, running behind the scenes to manage students' digital textbooks and analyze their performance' (Linday 2013). The IBM ideal of a 'classroom that will learn you' resonates with the notion that the smart city is an ambient intelligent environment that can 'think of us' as Crang and Graham (2007) have memorably phrased it. Its vision is of a classroom run on educational data scientific lines, with smarter learning software embedded into the pedagogic apparatus of the school.

However, in contrast to the image of the quantified student as a data object acted upon by algorithmic techniques, other smart city education projects feature a strong emphasis on the idea of 'smart citizens.' The basic logic is that the economic, cultural and political functioning of smart cities will rely on smart people that can help contribute to the monitoring and management of the city itself. Gabrys (2014: 38) for example argues that the citizen is increasingly viewed as a 'computational operative' in smart cities that are understood as 'datasets to be manipulated.' In this sense, smart cities appear to depend on citizens themselves becoming data practitioners.

One way in which smart citizens might be shaped as computational operatives of the smart city is by 'learning to code.' In recent years, initiatives designed to educate or train young people to learn programming skills have been proliferating, both in the UK and globally (Williamson 2015b). The organization Nesta in the UK has been a particular advocate of both learning to code and smart cities development, and has explicitly identified the contribution of the former to the latter, viewing learning to code as a kind of preparation for citizenship in a city where people are assumed to require the computational and data analytical skills to become operatives, engineers and hackers of the smart city's services and urban processes. Thus, in addition to sponsoring and supporting coding and 'digital making' initiatives for young people, such as Code Club and CoderDojo, through its Make Things Do Stuff campaign (Quinlan 2015), Nesta also supports 'civic technology' and 'coding for civic service' initiatives in the UK (Bell 2014), and Mulgan (2014), its chief executive, claims that it is 'promoting digital making of all kinds in cities,' particularly through its educational programs. Its manifesto *Rethinking Smart Cities from the Ground Up* describes how citizens might 'shape the future of their cities' through 'collaborative technologies', 'citizen sensing projects', and 'civic crowdfunding', and it promotes 'people-centred smart cities' which use 'open data and open platforms to mobilize collective knowledge', 'take human behaviour as seriously as technology', and 'invest in smart people, not just smart technology' (Saunders & Baeck 2015).

An accompanying Nesta report details how the city might act as a 'digital governor' to 'foster high-quality, low friction engagement with citizens'—by enabling citizens to interact with city services and input into urban policy making through digital interfaces—and to become a 'datavore' that turns big data into 'smart data' to 'optimize city services' by allowing citizens and businesses alike to access and build services from it (Gibson & Robinson 2015). The capacity of the smart city to become a 'digital governor' and a 'datavore' is, in Nesta's imaginary, dependent upon educating citizens to become digital producers and smart people; a task that is therefore delegated to programming clubs for young people and continued through civic coding projects where those individuals who have learned to code can contribute to the production of new digital interfaces to city services. A clear line is drawn in these Nesta documents and initiatives from learning to code and digital making to smart people-powered city governance.

In addition to this, Nesta has also been involved in projects to identify the 'data talent' required of the digital data sector in the UK in future years. It proposes that an 'Analytic Britain' will require new 'skills of the datavores,' and suggests that 'the pipeline of data talent starts in schools, where we need to ensure that the teaching of analytical skills is embedded across curricula,' including 'the use and impact of data, and data analytic and visualisation skills' (Mateos-Garcia, Bakhshi & Windsor 2015). An example provided in the report of how the new skills of datavores might be developed is Urban Data School (http://urbandataschool.org/), an educational initiative of the Milton Keynes smart city project MK:Smart. The aims of Urban Data School are to teach young people 'data literacy' to access and analyse urban datasets; create tools and resources to 'bring data skill education into the classroom'; to encourage new forms of 'active citizenship' through using data 'to design and evaluate Urban Innovation Projects'; and to devise 'effective solutions on the local, urban and global level.' Nesta's report on securing an analytic Britain recommends that 'The Government should keep a watching brief on innovative initiatives like the Urban Data School, and consider potential interventions to support their expansion where shown to have a beneficial impact' (Mateos-Garcia, Bakhshi & Windsor 2015).

Similarly, another example of how learning to code and smart cities are conjoined in emerging imaginaries of education is provided by the Future Makers program (http://open.glasgow.gov.uk/content/uploads/LiteracyEngagement.pdf), part of Glasgow's Future City initiative in the UK (a £24million government-funded smart cities showcase project). The Glasgow Future City vision emphasizes the 'literacies' required to 'empower and educate people in using city data' and the 'knowledge and skills to participate, understand or contribute to the Future City.' In order to promote these smart city literacies, the Future Makers program—which is facilitated by the Nesta-funded CoderDojo programming club—provides an 'innovative coding education programme' to develop programming and coding skills among young people. Future Makers consists of coding clubs and workshops all aimed at enabling young people to help shape and sustain the Future City. Related activities in the Glasgow Future City include 'Hacking the Future' events putting citizens, programmers, designers and government staff together in teams to focus on coding citizen-centred solutions to urban problems using the open datasets contained in the city's new 'data hub.' Future Makers thus acts in part to ensure young people are equipped with the relevant technical expertise of coding and data analysis to help 'hack' the future of the smart city.

In sum, through learning to code, young people are being trained as apprentice computational urbanists. As such, learning to code, digital making, data literacy and civic coding initiatives are all part of an emerging smart city vision which requires citizens to learn to code in order to help re-program, de-bug and optimize the software-supported city and all its urban services. As Vanolo (2014: 893) argues, 'citizens are very subtly asked to participate in the construction of smart cities' and 'implicitly considered responsible for this objective … in the form of an active citizen,' enabled to participate in the programming of apparently non-political solutions to problems of urban governance:

> In other words, citizens and local communities are invested with a moral obligation to behave in a certain way and adhere to the collective project of building smart cities; in this regard, the production of 'smart citizens' can be seen as an instrument of 'government at a distance.' (Vanolo 2014: 893-94)

In the smart city, the 'digital governor' acts at arm's length by subtly guiding the actions of citizens. In this context the actions of citizens have less to do with exercising rights, responsibilities and democratic engagement, and more with operationalizing computational processes 'so that smart cities will function optimally' (Gabrys 2014: 38).

As these examples indicate, the reimagining of education for the data-driven smart city runs along two interrelated lines of thought. First, smarter classrooms like those being designed by IBM are to become programmable educational spaces in which many aspects of administration, leadership, spatial organization, student management, communication and even pedagogy itself are to be governed through big data practices and processes. Smarter classrooms are in this sense distillations of the ambitions of educational data science. Second, new programs focusing on learning to code, data literacy and civic coding—such as those enacted by Nesta, Glasgow Future Makers, and MK:Smart Urban Data School—are positioning young people as apprentice data experts and computational urbanists. By equipping young people with the relevant data literacies and coding skills to enact appropriate data practices, these smart city initiatives seek to encourage them to occupy the forms of conduct that are appropriate for participation in smart cities, thus responsibilizing them as data analysts, digital makers and civic coders who will design the technologies that will enable the city, as a digital governor, to interact with its citizens. On the one hand, educating the smart city involves educating people to become smart citizens who can contribute to the design of digital urban infrastructures and devices, and in doing so contribute to an analytic Britain, and on the other it also involves the use of big data to enable the city itself to learn about all those individuals that inhabit it, and, as an increasingly smart learning environment, to reshape itself around their forms of behaviour and action.

Smarter startup schools

Though educational data science remains an emerging field of expertise, albeit one already infusing smart cities thinking, certain practices associated with an educational data science approach are already being inserted directly into some schools. IBM's P-TECH detailed above is one example, but others include AltSchool and Kahn Lab School, new institutions founded by Silicon Valley entrepreneurs and funded through Silicon Valley venture capital. These prototypical educational institutions originate in the ideals of Silicon Valley startup culture, and are designed to relocate its cultural practices to the whole system of schooling. Designed as scalable technical platforms and underpinned by software engineering expertise, they are also funded by commercial and venture capital and philanthropic sources; staffed and managed by entrepreneurs, executives and engineers from some of Silicon Valley's most successful startups and web companies; and proposed to reinvent, reimagine and rebuild education in the mould of Silicon Valley itself. In particular, they depend on a particularly enthusiastic big data imaginary associated with the culture of the technology entrepreneurship sector that sees data as the source of the solutions required to fix the 'broken' system of state schooling. Mager (2015: 5-6) describes 'algorithmic imaginaries' that emerge from 'a very specific economic and innovative culture' associated with Silicon Valley technology companies, and that privilege their originators' 'techno-euphoric interpretations of Internet technologies as driving forces for economic and social progress.' Silicon Valley's new startup schools are consistent with this notion of an algorithmic imaginary.

A prominent example is AltSchool (https://www.altschool.com/), set up in 2013 by Max Ventilla, a former tech entrepreneur and Google exec, which 'prepares students for the future through personalized learning experiences within micro-school communities.' Its stated aim is to 'help reinvent education from the ground up.' AltSchool originally raised $33million in venture capital funding, with a further $100million investment in 2015, including donations from Facebook's Mark Zuckerberg. Zuckerberg has subsequently announced plans for his own startup school, The Primary School (http://www.theprimaryschool.org/), to be launched in 2016. After establishing in four sites in San Francisco as a 'collaborative community of micro-schools,' AltSchool expanded in September 2015 to Brooklyn and Palo Alto, with further plans for new schools in 2016. It has since hired executives from Google and Uber plus other successful Silicon Valley startups. The AltSchool chief technology officer, formerly the engineer in charge of the Google.com homepage and search results experience, has stated that 'I am highly motivated to use my decade of Google experience to enable the AltSchool platform to grow and scale.' Elsewhere on the AltSchool site, the AltSchool 'platform' is described as a new 'central operating system for education,' a scalable technical infrastructure that can be transported to new sites, and it refers to 'technology-enabled models' that are transforming industries and institutions, such as Airbnb and Uber, and applies these ideals to education.

As a technical platform, AltSchool is managed on analytical, technical and scientific lines, albeit laced with the discourse of 'personalized learning' from which it draws its central philosophy—and which it shares with the wider educational data science field and IBM's Smarter Education program as shown earlier. As the AltSchool values claim:

> Our personalized learning approach puts each child at the center of everything we do … coupled with state-of-the-art classroom design and technology, [and] a flexible learning environment that mixes

> individual, group and experiential learning. … Our analytical approach and core strengths in innovation combine educational best practices with the latest tools. Our educators build learning experiences that are adaptive at their core and keep our children engaged.

Run on experiential learning principles, AltSchool encourages exploration, inquiry and problem-solving through the active construction of knowledge and understanding, whilst monitoring and regulating the experience through learning analytics and adaptive learning software.

AltSchool is, then, thoroughly governed, managed, and financed through the discourses and material practices of Silicon Valley startup culture. Its technical infrastructure as a platform is modelled on the big data practices associated with social media. Its funding is almost exclusively generated through venture capital. And its engineering and design team are applying their social media expertise in data analytics, algorithmic playlisting, adaptive recommender systems, and app development to the development of new ed-tech devices and platforms.

A similar arrangement is being developed by the founder of Khan Academy, the online platform that provides thousands of hours of 'practice exercises, instructional videos, and a personalized learning dashboard that empower learners to study at their own pace in and outside of the classroom.' Its entrepreneurial founder, Salman Khan, launched Khan Lab School (http://khanlabschool.org/) in September 2014. Located in Mountain View, in the San Francisco Bay Area, near Google HQ, Lab School is intended to realize the vision of schooling Kahn had previously outlined in his book *The One World Schoolhouse*. Kahn's Lab School teaches math, literacy and computer programming—in keeping with its tech sector roots—but like AltSchool also emphasizes 'real world' projects, personalized learning, student-centred learning, and a strong commitment to building children's 'character' and 'wellness' through, for example, 'mindfulness' meditation training.

Most notably, however, Kahn Lab School has been established as an experimental R&D lab for testing out different educational approaches and technologies, and aspires to the production of new theories of learning itself. Lab School has been profiled in *Wired*, which noted its

> goal isn't just to build one fancy school but to develop and test a new model of learning that can be exported to other schools around the country and the world. His team is diligently recording and tracking every student's progress and sharing the findings with their parents and the staff, an open source approach to educational innovation. In this view, the Lab School kids are guinea pigs … willingly subjecting themselves to new ideas that have never been tried before, then adapting and adjusting and trying again. … 'This is a lab for establishing new theories that could affect the rest of the planet,' Khan says. (Tanz 2015)

Like AltSchool above, Lab School's 'touchy-feely surface' of character education 'masks a rigorous fealty to tracking data about every dimension of a student's scholastic and social progress' (Tanz 2015). In particular, it uses data analytics to provide a constant and growing trace of the character development of its pupils, and reinforces those data through standardized testing. It also welcomes outside organizations in to the school to test out new ideas and technologies, so that the children are positioned as constant subjects of a tech experimentalist approach. AltSchool likewise engages its students in regular HCI-based experiments to develop, test and finetune its operating system. In short, startup schools such as AltSchool and Kahn Lab School are set up as experimental testbeds for smarter learning software systems, and position schools as analytics laboratories that treat their students as sources for data extraction, aggregation and pattern-detection; data that can then be used as indicators of character and used as insights to inform new character interventions.

The data practices associated with startup schools raise the particular issue that as they are privately owned and governed, any student data generated by these institutions also belongs to them to use to conduct various analytics procedures. This circumvents wider concerns among parents and critics about access to student data by third party commercial organizations. Instead, these startup schools have direct access to their data, and can collect and calculate it in-house—a significant example of the 'capture model' (Agre 1994) of data collection that allows computers to track information in real time, identify particular human activities, and reorganize the data sets in ways that can be used for intervention. In this way startup schools act much like the social media companies from which they are derived, whose business plans depend on the capture and analysis of customer and user data, often with little external scrutiny, for the purposes of better profiling and prediction of individuals' habits and social trends. The implantation of data analytics in the everyday functioning of schools by key figureheads of these companies—tracking and predicting everything from academic attainment to behavioural 'character' indicators and emotions in classrooms—should itself be the topic of close scrutiny. How might startup schools collect data, who will own it, what will they do with it, and to what ethical codes will they subscribe? While prototypical startup schools may be conceived as 'angel investments' from Silicon Valley to the social institution of education, they also need to be understood as the products of an imaginary that is rooted in a distinctive techno-euphoric culture that emphasizes technical innovation and hacking, commercial business planning, and social media data capture. The glossy imaginary of smarter startup schools conceals how they are also surveillant, data-capturing, experimental laboratories and scalable venture capitalist enterprises built to run on the social, cultural, economic and political operating systems of Silicon Valley itself.

Conclusion: Data and algorithms of the powerful

This paper has surveyed some key features of a socially shared big data imaginary that is beginning to animate the development of a range of technical projects in education. It has detailed how a first wave of big data in education emerged through the exhibitionary practices of the nineteenth-century great expositions, and how practices of data collection, calculation and communication have been continued in large-scale educational databanks such as the National Pupil Database and Pearson's Learning Curve. It has traced the formation of a new field of educational data science, an interdisciplinary hybrid of methods originating in computer science and psychological learning sciences (CompPsy) that aims to produce a big data infrastructure for education and elicit from the data any patterns that might indicate new theories of learning itself. As the idea of data-driven 'smart cities' has gained popularity, educational institutions have become the focus of intense re-imaginings that see them as data platforms in which learners are subject to real-time predictive analytics and adaptive learning platforms, or which treat young people as apprentice data analysts who should learn to code and to analyse data in order to help 'hack' the future of their cities and contribute to the 'data talent' pipeline in the UK. Finally, the paper has identified the new Silicon Valley startup schools where the big data imaginary associated with the culture of the technology entrepreneurship sector is being materialized directly as a new central operating system for schools.

These constitute a wide variety of practices and ambitions, but my central contention throughout has been that they are all animated by an increasingly shared imaginary of the future of education that its advocates believe is desirable and attainable through the application of technologies and practices associated with big data. The specific big data imaginary of education detailed in this paper is beginning to stimulate new kinds of practices that are intended to transform educational institutions, processes and practices in a variety of ways: by reshaping policies, by articulating new theories, by galvanizing new adaptive learning platforms that can respond automatically to interaction with learners, by sculpting young people as data practitioners, and by remaking schools according to the data-driven culture of technology entrepreneurs. Through these processes and practices, the big data imaginary of education is becoming a smart educational reality to be inhabited, an emerging educational 'species' of big data (Kitchin & McArdle 2015).

Etymologically, 'smart' refers both to education (to become smart) and to software (as in software smarts, smartphones, smart cities and so on). The notion of smarter education, running on smart learning software, raises a significant final issue. Smart learning software itself has to be trained, or educated, to perform as planned. Its underlying processes such as machine learning rely on adaptive algorithms and statistical models that need to be 'fed training data'; these are, crudely speaking, 'taught algorithms' that can learn from being taught with example data, but that sometimes fail to perform as expected 'in the wild' (Gillespie 2014). As Mackenzie (2015) notes, machine learning algorithms have to be constantly retrained in an iterative process of monitoring, adjusting, revising and optimizing as the accuracy and generalizability of the predictive models they generate are themselves checked and analysed. Very little is known, however, about the ways in which smart learning software systems such as learning analytics and adaptive learning platforms like those being deployed by Knewton and Pearson and IBM are being trained and retrained to become smart. What training data are these systems learning from? Who has selected those data, on what basis, and for what purpose? What patterns will they look for in the data, which will they value, what will they identify as normal or deviant through techniques such as 'algorithmic assessment-norming at scale,' as Knewton describes it?

These questions reflect long-standing questions in educational sociology about the selection of knowledge for inclusion in school curricula, and claims that schooling tends to reinforce and reproduce the interests and 'knowledge of the powerful.' In relation to IBM's smarter classrooms, Pearson, AltSchool, Knewton and the rest, we might need to ask questions about how the 'data and algorithms of the powerful' are now being used to train the smart learning software that is being proposed to augment and optimize the classroom of the future. How the training data is prepared, or how the data analytics and pattern recognition techniques are designed, are deeply consequential to how software will learn from the subjects they are applied to, and how they might then adapt and respond to them pedagogically. Educational data scientists, learning analytics providers and related experts have become powerful educational actors, mobilizing big data practices to reconfigure classrooms as smart learning environments that depend on the judgments and values of their designers. Much more needs to be done to explore how the functioning of smart future educational institutions may be governed by the production and selection of the data and algorithms of the powerful, and how they might act to make future educational realities conform to the imaginaries of their producers.

Acknowledgement

This work was supported by a research grant from the Economic and Social Research Council (ref: ES/L001160/1): https://codeactsineducation.wordpress.com/

References

Agre, P. 1994. Surveillance and capture: Two models of privacy. *Information Society* 10, no. 2: 101-127.

Ambrose, M.L. 2014. Lessons from the avalanche of numbers: big data in historical perspective. Available at: http://ssrn.com/abstract=2486981

Batty, M. 2013. *The New Science of Cities*. London: MIT Press.

Batty, M., Axhausen, K.W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. 2012. Smart cities of the future. *Eur. Phys. J. Special Topics* 214, 481–518.

Behrens, J. 2013. Harnessing the Currents of the Digital Ocean. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA., April, 2013.

Bell, H. 2014. Coding for civic service: what we are learning? Nesta blogs. Available online: http://www.nesta.org.uk/blog/coding-civic-service-what-we-are-learning

Buckingham Shum, S., Baker, R.S.J., Behrens, J., Hawksey, M., Jeffery, N. & Pea, R. 2013. Educational data scientists: a scarce breed? LAK '13, 8-12 April 2013, Leuven, Belgium: http://simon.buckinghamshum.net/wp-content/uploads/2013/03/LAK13Panel-EducDataScientists.pdf

Crang, M. & Graham, S. 2007. Sentient cities: ambient intelligence and the politics of urban space. *Information, Communication & Society* 10, no. 6: 789-817.

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

Dourish, P. & Bell, G. 2011. *Divining a Digital Future: Mess and Mythology in Ubiquitous Computing*. London: MIT Press.

Gabrys, J. 2014. Programming environments: environmentality and citizen sensing in the smart city. *Environment & Planning D: Society & Space* 32: 30-48.

Gibson, J. & Robinson, M. 2015. *CITIE: City Initiatives for Technology, Innovation and Entrepreneurship*. London: Nesta.

Gillespie, T. 2014. Algorithm. *Culture Digitally*, 25 June: http://culturedigitally.org/2014/06/algorithm-draft-digitalkeyword/

Hacking, I. 1990. *The Taming of Chance*. Cambridge: Cambridge University Press.

Hogan, A., Sellar, S. & Lingard, B. 2015. Commercialising comparison: Pearson puts the TLC in soft capitalism. *Journal of Education Policy*. http://dx.doi.org/10.1080/02680939.2015.1112922

Hope, A. 2014. Foucault's toolbox: critical insights for education and technology researchers. *Learning, Media & Technology* 40, no. 4: 536-549.

Housley, W. 2015. The emerging contours of data science. *Discover Society* 23, 3 August 2015: http://discoversociety.org/2015/08/03/focus-the-emerging-contours-of-data-science/

Jasanoff, S. 2015. Future imperfect: Science, Technology, and the Imaginations of Modernity. In S. Jasanoff & S.-H. Kim (eds). *Dreamscapes of Modernity: Sociotechnical Imaginaries and the Fabrication of Power.* Chicago: University of Chicago Press.

Kahn. S. 2012. *The One World Schoolhouse: Education Reimagined*. New York: Twelve/Hachette Book Group.

Kitchin, R. 2014. The real-time city? Big data and smart urbanism. *GeoJournal* 79: 1-14.

Kitchin, R. & McArdle, G. The diverse nature of big data. *The Programmable City Working Paper 15*. Available at: http://ssrn.com/abstract=2662462

Lawn, M. 2013. The rise of data in education. In M. Lawn (ed.) *The Rise of Data in Education Systems: Collection, visualization and use*: 7-10. Oxford: Symposium.

Linday, G. 2013. IBM's Department of Education: The company that brought you smarter cities moves into schools. *NextCity*, 19 August 2013: http://nextcity.org/features/view/ibms-department-of-education

Mackenzie, A. 2015a. The production of prediction: What does machine learning want? *European Journal of Cultural Studies* 18, no. 4-5: 429-445.

Mager, A. 2015. Glocal Search: Search technology at the intersection of global capitalism and local socio-political cultures. Vienna: Institute of Technology Assessment (ITA), Austrian Academy of Sciences. Available at: http://www.astridmager.net/wp-content/uploads/2015/11/Abschlussbericht-OeNB_Mager.pdf

Mateos-Garcia, J., Bakhshi H., & Windsor, G. 2015. *Skills of the Datavores: Talent and the data revolution*. London: Nesta.

Mayer-Schönberger, V. & Cukier, K. 2014. *Learning from Big Data: The Future of Education.* New York: Houghton Mifflin Harcourt Publishing Co.

Mulgan, G. 2014. Smart cities: what we're doing and why. Nesta blog, 23 January: http://www.nesta.org.uk/blog/smart-cities-what-were-doing-and-why

Pea, R. 2014. A report on building the field of learning analytics for personalized learning at scale. Stanford: Stanford University.

Piety, P.J., Behrens, J. & Pea, R. 2013. Educational data sciences and the need for interpretive skills. American Educational Research Association, 27 April-1 May 2013, San Francisco: http://www.slideshare.net/PhilipPiety/education-data-sciences-and-interpretive-skills

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, IN, USA: http://edinfoconnections.com/wp-content/uploads/2014/01/Educational-Data-Sciences-Feb-9.pdf

Quinlan, O. 2015. *Young Digital Makers: Surveying attitudes and opportunities for digital creativity across the UK.* London: Nesta.

Robertson, H. & Travaglia, J. 2015. Big data problems we face today can be traced to the social ordering practices of the 19th century. *LSE Impact of Social Sciences*, 13 October 2015: http://blogs.lse.ac.uk/impactofsocialsciences/2015/10/13/ideological-inheritances-in-the-data-revolution/

Ruppert, E. *et al.* 2015. Socializing big data: from concept to practice. CRESC Working Paper no. 138: http://www.cresc.ac.uk/medialibrary/workingpapers/wp138.pdf

Saunders, T. & Baeck, P. 2015. *Rethinking Smart Cities from the Ground Up*. London: Nesta.

Shepard, M. 2011. Toward the sentient city. In M. Shepard (ed.) *Sentient City: ubiquitous computing, architecture, and the future of urban space:* 10-35. Cambridge, MA: MIT Press.

Sobe, N. 2013. Educational data at late nineteenth- and early twentieth-century international expositions: 'accomplished results' and 'instruments and apparatuses.' In M. Lawn (ed.) *The Rise of Data in Education Systems: Collection, visualization and use*: 41-56. Oxford: Symposium.

Tanz, J. 2015. The Tech Elite's Quest to Reinvent School in Its Own Image. *Wired*, 26 October 2015: http://www.wired.com/2015/10/salman-khan-academy-lab-school-reinventing-classrooms/

Thrift, N. 2014. The sentient city and what it may portend. *Big Data & Society* 1, no. 1: http://dx.doi.org/10.1177/2053951714532241

Townsend, A.M. 2013. *Smart Cities: Big data, civic hackers and the quest for a new utopia*. London: WW Norton & Co.

Vanolo, A. 2013. Smartmentality: The smart city as disciplinary strategy. *Urban Studies* 51 no. 5: 883-89.

Verebes, T. (ed.) 2014. *Masterplanning the Adaptive City: Computational urbanism in the twenty-first century*. London: Routledge.

Williamson, B. 2015a. Digital education governance: data visualization, predictive analytics and 'real-time' policy instruments. *Journal of Education Policy*. http://dx.doi.org/10.1080/02680939.2015.1035758

Williamson, B. 2015b. Political computational thinking: policy networks, digital governance, and 'learning to code'. *Critical Policy Studies*. http://dx.doi.org/10.1080/19460171.2015.1052003

Williamson, B. 2015c. Educating the smart city: schooling smart citizens through computational urbanism. *Big Data & Society* 2, no.2: 1-13. http://dx.doi.org/10.1177/2053951715617783

Williamson, B. 2016 *forthcoming*. Digital methodologies of education governance: Pearson plc and the remediation of methods. *European Educational Research Journal*. http://dx.doi.org/10.1177/1474904115612485