EERJ European Educational Research Journal

Digital methodologies of education governance: Pearson plc and the remediation of methods

Ben Williamson
School of Education, University of Stirling, UK

Abstract
This article analyses the rise of software systems in education governance, focusing on digital methods in the collection, calculation and circulation of educational data. It examines how software-mediated methods intervene in the ways educational institutions and actors are seen, known and acted upon through an analysis of the methodological complex of Pearson Education's Learning Curve data-bank and its Center for Digital Data, Analytics and Adaptive Learning. This calls for critical attention to the 'social life' of its methods in terms of their historical, technical and methodological provenance; their affordances to generate data for circulation within the institutional circuitry of Pearson and to its wider social networks; their capacity to configure research users' interpretations; and their generativity to produce the knowledge to influence education policy decisions and pedagogic practices. The purpose of the article is to critically survey the digital methods being mobilized by Pearson to generate educational data, and to examine how its methodological complex acts to produce a new data-based knowledge infrastructure for education. The consequence of this shift to data-based forms of digital education governance by Pearson is a challenge to the legitimacy of the social sciences in the theorization and understanding of learning, and its displacement to the authority of the data sciences.

Emerging digital methods of data collection, calculation and communication are intervening in how educational institutions and actors are seen, known and acted upon. This article provides an analysis of the methodologies of Pearson plc's Learning Curve data-bank and its Center for Digital Data, Analytics and Adaptive Learning. Pearson's mobilization of digital methods exemplifies a

Corresponding author:
Ben Williamson, School of Education, University of Stirling, Pathfoot, Stirling FK9 4DS, UK.
Email: ben.williamson@stir.ac.uk

shift towards more software-based, computer-coded and algorithmically mediated techniques of educational governance, and is integral to the production and performativity of a powerful educational data infrastructure. Pearson is examined as an important actor with the relevant digital methods of software-mediated statistical data collection, analysis, visualization and interactivity to make contemporary education legible, intelligible, and therefore actionable through highly targeted intervention. Its Learning Curve enables this through providing statistical indices and data visualizations representing the global comparison of national education systems, while its Digital Data, Analytics and Adaptive Learning centre focuses on mining patterns from individual learners' data using advanced data analytics software in order to derive new insights into the learning process itself and then design new e-learning products based on those insights.

These are significant methodological accomplishments, and this article interrogates the 'social life of methods' (Savage, 2013) deployed by Pearson—the provenance of such methods, and the ways they are inscribed with particular values and assumptions that then shape the insights they generate. The consequence of this shift to data-based methods of 'digital education governance' (Williamson, 2015) by Pearson is a displacement of the legitimacy of the social sciences in the theorization and understanding of learning to the authority of the data sciences. The 'psy complex' of the psychological sciences (Rose, 1999a) dominated 20th-century attempts to model and classify the learning process by implanting the gaze of 'psychological eyes' in education (Popkewitz, 2012). For Pearson it is the computational affordances of data science in combination with learning science methods and theories that promise to produce the classifications and models by which learning is to be understood and acted upon in the data scientific 21st century. The combination of the methods of learning sciences with computer-based forms of data science is embryonic of an evolving 'CompPsy' complex that hybridizes theories, concepts and practices from the computer sciences (CompSci) with those of the psy-sciences (Loveless and Williamson, 2013). The central aim of the article is to trace the consequences emerging from the digital methods that Pearson is employing in the classification and modelling of learning for the ways in which education might be governed.

Socializing methods

Across the social sciences, the production of social and human knowledge and theory is being affected by digital data analyses enacted through software (Kitchin, 2014). The emergence of big data software and its algorithmic techniques of analysis is challenging conventional views about the institutional practices and spaces of knowledge production (Burrows and Savage, 2014). Instead of social scientists, it is claimed, the new experts of the social world are 'algorithmists'— experts in computer science, mathematics and statistics, as well as policy, law, economics and social research—who can undertake big data analyses across commercial, political and scholarly sites (Mayer-Schonberger and Cukier, 2013). While the 'everyday user performs their social life via Facebook, Twitter et al.', these experts apply their 'methodological techniques for spotting the movement of trends and behaviours' (Davies, 2015: 13).

The emergence of 'digital sociology' and 'digital social research' reflects disciplinary anxieties about the relevance of social science at a time when commercial social media companies, research and development (R&D) labs and think tanks are staking their claim to methodological expertise (Lupton, 2015). 'Digital methods' that involve the use of digital devices to 'perform a cultural and social diagnostics' (Rogers, 2013: 3) have the capacity to detect patterns in huge quantities of data and to augment how people and societies see and know themselves. As a consequence, there has been a 'redistribution of social research' between human actors—researchers, software developers, data analysts, commercial R&D labs and social media users—but also to non-human actors

including databases, software, algorithms, platforms and other digital devices (Marres, 2012). The redistribution of research also entails a remediation of methods as social research is refashioned through digital data-processing technologies. The emergence of a new field of 'educational data science' (Piety et al., 2014) reflects how digital methodologies have been distributed into educational research too.

Recent research on 'the social life of methods' has engaged with the plethora of digitally remediated methods for performing social scientific research. By social life of methods, what is meant is a critical engagement with methodological devices that resists framing them simply as technical tools, but makes their affordances and capacities into the object of social scientific inquiry (Savage, 2013). Firstly, methods are social because they are shaped by the social, cultural, economic and political circumstances in which they have been produced and of which they are a part (Law et al., 2011). Methods are designed for particular purposes, through the work of advocates, as devices for examining, seeing, knowing and interpreting the social world. Secondly, methods are also social because they in turn help to shape that social world. The discoveries made by social scientific research conducted through digital devices, then, are not objective facsimiles of an existing world, but are consequential for 'social scientific ways of knowing' (Ruppert et al., 2013: 24). As part of the material of contemporary ways of life, digital devices are fundamentally reconfiguring the ways in which social science can be performed, and the kinds of analyses, interpretations and insights into social worlds made possible. Digital data are therefore socially, technically and methodologically enacted 'data practices' with their own social lives, and the *generation* of data is therefore *generative* of particular effects and social implications; data are consequential to 'what is known,' and can influence decision-making and other activities (Ruppert et al., 2015). The notion of 'socializing methods' registers this double process of socially enacting methods, and of mobilizing methods to make sense of the social world and wrap new social norms around it.

Drawing on ideas about remediating and socializing methods as an analytical framework, I argue Pearson has become a significant methodological gatekeeper in the mobilization of digital methods in education, thus in defining how and what can be seen and known about it, and what consequently might be done to govern it. Methodological expertise in data science is increasingly being redistributed to organizations such as Pearson, and remediated through their digital methods, rather than enacted within either the methodological and theoretical apparatuses of research departments in universities or the data analysis agencies associated with government departments. Pearson's current efforts around digital methods represent the embedding of a new form of classification in the knowledge infrastructure of education. Whereas the process of learning was, throughout the 20th century, largely the expert preserve of the psychological sciences, which translated 'the discourse of science to the imperatives and techniques of practice' (Rose, 1999a: 201) and acted as a relay between psychological models of development and the practices of the school, for Pearson the process of learning is now to be mapped and known through the data sciences, albeit twinned with learning science insights from cognitive and developmental psychology, neuroscience and behavioural science. Pearson's hybrid science of data and learning, and the methods that enact it, are consequential to ways of conceptualizing learning processes, measuring learner progression and developing pedagogic products and practices. Through its methodological complex of psychological and data scientific ways of knowing and intervening in learning processes, Pearson is seeking to derive new classifications and standard definitions of learning itself that can then be relayed into practices by being coded into the e-learning software products that it inserts into the pedagogic routines of the classroom.

As Bowker and Star (1999: 314) have argued, classifications and categories 'touch people in a variety of ways—they are assigned, they become self-chosen labels, they may be statistical artefacts'. The new digital methods and data practices enacted by Pearson are becoming active in the

production of classificatory categories in relation to learning that have the potential to touch people's lives by rendering new models and understandings of what constitutes learning itself. Pearson is seeking to embed such classifications in its recommendations for a new form of educational governance that focuses on personalizing the learning process—a task to be enacted by new data-producing and -processing devices embedded in the pedagogic structure of the school—rather than solely through the bureaucracy of education systems. In this sense, Pearson's digital methods are key techniques in the generation of a data-derived classification system for learning, and are consequential to the production of new knowledge in relation to the ways that learning is known and learners themselves are made amenable to being acted upon pedagogically. This reflects a structural shift in the system of education governance from centralized bureaucracies to non-state and non-public sector organizations (including commercial companies), and a discursive shift from education to learning (Ozga et al., 2011). It is also part of a shift to focus on the subjectification of individuals through diverse practices and 'technologies of schooling' that are intended to shape their capacities for thought and action (Rose, 1999b: 54).

The soft(ware) governing techniques of Pearson plc

Pearson plc is the world's largest educational publisher. Originally established in 1844, it announced $7.9bn USD revenue in 2014 with operations in over 70 countries and over 40,000 employees (https://www.pearson.com/about-us.html). Following a re-structuring and re-branding exercise in 2014, represented by its strapline 'always learning' and its goal 'to help people make measurable progress in their lives through learning', Pearson has significantly broadened its field of operations to include major digital platforms for online publishing, testing and assessment, data analysis and digital research, and has established an 'affordable learning fund', a free-enterprise model of low-cost private schools for low-income countries (Ball and Junemann, 2015). In 2014 it also successfully tendered to provide the frameworks for the Organization for Economic Cooperation and Development (OECD) Programme for International Student Assessment (PISA) tests scheduled for 2018; the frameworks define what will be measured in the test instruments, how this will be reported and which approach will be chosen for the development of tests and questionnaires (Pearson, 2014).

Pearson has therefore not only transformed itself from 'a media holding company to an edu-business', but also positioned itself as a 'legitimate policy actor' and a 'morally authoritative agency in educational matters' (Hogan et al., 2015: 49). It has also committed to measuring the learning outcomes of its products and services in order to enable the company to demonstrate the extent to which any Pearson product has a measurable impact on improving the user's life through learning. As part of this, it has established an 'Open Ideas' database of reports 'to help make the best evidence and ideas about learning accessible to all, and to encourage open debate about what works in education' (https://research.pearson.com/). In 2015 it launched two reports, one entitled 'What Works in Education' and the other 'What Doesn't Work in Education' (http://blog.pearson.com/what-works-in-education-a-tough-love-message-from-john-hattie/), and it has also established an Efficacy Framework, a 'tool that uses a tried and tested method to help understand how products or services can achieve their intended outcomes or results' (http://efficacy.pearson.com/efficacy-tool/). As a global 'edu-business' with links to government, commercial and multilateral agencies, Pearson has become a 'serious policy player' that can both define problems and solve them but often 'goes unnoticed in education policy analysis' (Ball, 2012: 128).

In this context, I analyse the role of two of Pearson's recent developments in identifying both policy problems and solutions, focusing especially on its digitally mediated data-processing methodologies. These are the Learning Curve, a global databank and source of analysis on education

launched in 2012 (http://thelearningcurve.pearson.com/), and its Center for Digital Data, Analytics and Adaptive Learning (CDDAAL), a R&D centre dedicated to the analysis and use of digital data for educational improvement also established in 2012 as one of five centres in Pearson's Research and Innovation Network (http://researchnetwork.pearson.com/digital-data-analytics-and-adaptive-learning). In combination, and supported by other Pearson documents, these resources emphasize Pearson's transformation into a company that develops digital learning resources for use in schools, the data-processing technologies and methods for analysing the data produced by them, and also the data analytics and visualization tools required to measure and monitor the efficacy of whole education systems.

Recent studies have traced Pearson's networks of influence into the educational policy sphere (Hogan et al., 2015), and emphasized 'Pearson's overall business ambitions … to find new markets and to create new spaces of education for Pearson's products' (Ball and Junemann, 2015: 49). The policy networks and commercial ambitions of Pearson are part of the argument in this article, but its novel claim is that Pearson acts as a global methodological gatekeeper in defining and modelling what constitutes learning, and that this act of classification is consequential to how education systems and individual learners alike will be governed in the future. Beyond its business plans, Pearson is participating in a reconfiguration of the methods by which learning is conceptualized, measured and understood, and seeking to secure consensus for its views through mobilizing techniques of data visualization and human–computer interaction (HCI). The aim of the article is to examine the scope of Pearson's methods to produce and circulate knowledge about learning, and the argument is that such knowledge may be redefining existing conceptualizations about learning and its measurement that have previously been the preserve of social scientific forms of expertise. Pearson is not merely seeking new market niches, but redefining learning itself and seeking to mobilize its knowledge about learning and cognition in the specification of new pedagogic applications and products. Focusing on the specific methodological and technical instruments it mobilizes is crucial to understanding how its policy and commercial ambitions are being operationalized and how its goal 'to help people make measurable progress in their lives through learning' is materialized.

As such, the role of Pearson in influencing policy processes and pedagogic practice is part of a wider 'governance turn' in European education and beyond (Ozga et al., 2011). Increasingly, governance is conceived as a form of 'soft power' realized through techniques of attraction, seduction, persuasion and the cultivation of support and shared interest across networks of loosely associated actors from across the public and private sectors (Moos, 2009), including 'those conventionally considered peripheral to education governance' such as 'commercial interests and technological innovators' (Lawn and Grek, 2012: 82). The shift to soft governance is enabling a new kind of governing expert whose claim to authority rests on the methodological and technical capacity to know, assess and act upon education through data collection, aggregation and analysis, and to produce new kinds of 'governing knowledge' (Fenwick et al., 2014). Websites and online portals that present the data persuasively as the knowledge required to facilitate governing practices, and the data practices employed to generate them, have therefore become the focus of recent research (e.g. Decuypere et al., 2014; Piattoeva, 2015; Williamson, 2015). Following a long historical rise in the use of data in education systems (Lawn, 2013), such studies demonstrate how education governance is currently being accelerated by software-mediated processes of 'datafication', including:

> …the conceptualisation and codification by which the pre-existing frames, categories and classifications shape the information that is constituted as data and which influence the possibilities for its usage and effects …; the algorithmic treatment of data through which patterns and correlations are produced; and the re-representation of the world through data visualisation and the navigation of data by users. (Sellar, 2015: 132–133)

The 'data infrastructure' underpinning the production of governing knowledge is a sophisticated technical and methodological accomplishment. As Sellar (2014: 6) argues, the concept of data infrastructure in relation to education governance can be defined as 'an assemblage of material, semiotic and social practices' that functions to translate things into numbers; enables the storage, transmission, analysis and representation of data using algorithmic logics and computational technologies; embeds data usage into other practices; and produces new kinds of spaces and social practices through practices of classification, measurement and comparison. Likewise, Edwards et al. (2013: 5) refer to 'knowledge infrastructures' as 'networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds'. It is useful to think of education as being orchestrated through infrastructures in which digital data practices are now embedded in the production of new governing knowledge about educational institutions, practices and spaces, and where data are increasingly perceived and accepted as a form of authoritative, objective and impartial knowledge. Yet, while the 'social technologies' of soft governance appear to be 'natural' or 'neutral' tools (Moos, 2009), they are deeply inscribed with the methodological values and styles of thinking of their designers and sponsors.

Methodologically, I examine the Learning Curve and CDDAAL resources through detailed documentary analysis of the various websites, reports and visualizations produced to support them, focusing explicitly on the digital methods involved in the generation of the Learning Curve and enacted by the CDDAAL. These methods constitute what I term Pearson's methodological complex—the technical and practical instruments for knowing and classifying learning. Drawing on sociomaterialist approaches from science and technology studies (STSs), and consistent with the social life of methods approach, a methodological complex does not only consist of the technical instruments themselves, but a sociotechnical apparatus including the human and social actors designing and deploying them; the institutions promoting and sponsoring their use; the epistemological assumptions underpinning them; the underlying software, code and algorithms that enable them to function; assumptions about the users for whom they are intended to produce data; and the productivity of such an apparatus to exert material effects by shaping social practices and influencing decision-making. This ultimately represents what Bowker and Star (1999) have influentially designated as an infrastructural system of classifications, standards and categories that loops back into the social world it represents. Pearson enforces a particular set of classifications and models of what learning is—as defined through its hybrid data science/learning science methodologies—which then informs policies, decisions and technical interventions that constitute those understandings in particular material practices. This is a knowledge infrastructure consisting of data collected through learning processes; that produces knowledge about learning from analyses of these data; and that is consequential to ways of intervening in future learning processes and learners' lives. Such an infrastructure is the material instantiation of Pearson's techniques of soft governance: a system delegated to non-government experts, with associated techniques that seek to activate the capacities of subjects to act in new ways.

Digital methods

Pearson does not possess the direct means to set education policy or determine pedagogy through hard regulative governance. Instead, it governs more softly and indirectly through seeking to attract policymakers, practitioners and other publics to the insights, recommendations and advice it is able to derive from data. In this section I critically examine six categories of digital methods being developed and deployed by Pearson, specifically analysing its role as a methodological gatekeeper with the technical expertise to produce new classifications, categories and models of learning and to construct a new data/knowledge infrastructure for education, understandings that it can then

codify into software products that can be inserted into the pedagogic techniques and technologies of the classroom.

Statistical methods

Much has been written in the field of education policy research on the process of 'governing through data' (e.g. Gorur, 2013; Grek, 2009; Ozga et al., 2011), whereby numbers and a 'logic of enumeration' (Hardy, 2015) are used to produce the knowledge required to enable education to become governable. Such studies contend that numbers are never merely factual, transparent or theory-free conveyors of reality, but the product of particular languages, categories, interpretations and doctrines that result in the production of norms and expectations—such as what should happen in schools. 'Commensuration' has become a particularly productive statistical method in the transformation and standardization of different qualities into a common metric as 'evidence' required by policymakers for benchmarking, comparison, evaluation and decision-making purposes (Sellar, 2015).

Pearson's Learning Curve exemplifies the productivity of numbers to influence policy decisions. Launched in November 2012 under the leadership of Michael Barber (Pearson's chief education adviser, and a former head of the UK Prime Minister's Delivery Unit; see Barber and Ozga, 2014), the Learning Curve consists of a vast databank of educational data aggregated together from over 60 datasets from around the globe. According to its website:

> Through The Learning Curve we are contributing to the global conversation on learning outcomes; to help positively influence education policy at local, regional and national levels. The data and analysis on this website will help governments, teachers and learners identify the common elements of effective education.

The data in the Learning Curve databank have been compiled into a global index of educational performance that maps correlations between the inputs to and outputs of education, the inputs to education and socio-economic environment indicators (as a proxy for wider society), and the outputs of education and socio-economic environment indicators. These data are presented on the website as ranked league tables, visual tools and also compiled into reports (to date reports are available from 2012 and 2014). In his foreword to the 2012 report, Michael Barber wrote that the Learning Curve would become 'an open, living database which we hope will encourage new research and ultimately enable improved … evidence- informed education policy'.

Underpinning the Learning Curve is a complex of statistical methods utilized to ensure the commensurability and comparability of the data from the different datasets. As stated in the methodology appendix to the 2012 report:

> The aim of the Data Bank was to include only internationally comparable data. Wherever possible, OECD data or data from international organisations was used to ensure comparability … and when possible, used inter- and extrapolations in order to fill missing data points. Different methods for estimations were used, including regression when found to be statistically significant, linear estimation, averages between regions, and deductions based on other research.

Elsewhere in the appendix, it is possible to identify a number of methodological commitments and epistemological assumptions. It refers to 'objective quantitative indicators' and the normalization of statistical indicators into 'z-scores' to indicate how many standard deviations an observation is above or below the mean. Notably, the appendix features a number of cautions, such as that 'because indexes aggregate different datasets on different scales from different sources, building

them invariably requires making a number of subjective decisions'. This is commensuration materialized methodologically: a common metric derived from the transformation of different datasets and qualities into standardized form. It also registers what actor-network theorists have termed 'qualculation'—how data are made to qualify for inclusion in calculations (see Edwards and Fenwick, this issue).

From a socializing methods perspective, it is especially significant that the Learning Curve is the product of the Economist Intelligence Unit (EIU) (http://www.eiu.com/). Pearson itself owns 50% of the Economist Group of which the EIU is the research and analysis division. The EIU field of expertise is in economic and market data; its 'sound and transparent' methodologies include economic, political and socio-demographic forecasting; quantitative, qualitative and synthetic indicators; innovative scoring systems; statistical index construction and global ranking; multi-dimensional comparison; data modelling and scenario analysis; industry and risk analyses; and economic, political, cultural and locational benchmarking. The EIU enacts these methods in research for business as well as government. For the latter, its 'team of analysts, economists, and regulatory specialists' are ' helping clients develop data-driven solutions to public policy challenges', and it claims its 'research programmes, always supported by reliable data and actionable results, have helped governments, foundations, NGOs and business associations to understand and overcome the challenges they face in the world of public policy'. Many of these methods are remobilized in the Learning Curve, in its global index of ranked countries; its production of country profiles detailing their social, political and demographic indicators; its generation of visualizations to make the data easy to use and interpret; and its attempt to correlate the inputs to and outputs of education with socio-economic environment indicators. These methods are drawn from the repertoire of business and market intelligence, and infuse educational data with economic logics of calculation and forecasting. In particular, the EIU's expertise in country comparison and global benchmarking infuses the design of the Learning Curve to promote cross-country comparison of education systems.

Its statistical methods permit Pearson to position itself as a 'centre of calculation' in the governance of education, enabling it 'to act as a centre by means of its centrality in the flows of information that "re-present" that over which it is to calculate and seek to programme' (Rose, 1999b: 211). By turning education into numericized inscriptions, Pearson's methods render it visible as a calculable space defined according to specific statistical methods and the norms, epistemologies and assumptions underlying them. There is a double social life to Pearson's statistical methods, in that the data so generated do not merely inscribe a pre-existing reality but constitute education as a calculable space in which institutions (schools, local authorities, government education departments) are incited to calculate about themselves in certain ways, and act to improve and optimize themselves 'because they are calculated about in certain ways *by others*' (Rose, 1999b: 213). The social life of the EIU's methods is, in other words, consequential to the ways in which education is centred for counting and calculation in the Learning Curve.

Data science methods

While statistical methods are a dominant aspect of the Learning Curve, Pearson is also developing more digitally native methods from the field of data science to support its production of new knowledge about learning and skills. Building on established statistical methods and models, data analytics technologies utilize advances in information management and storage, data handling, modelling algorithms, machine intelligence and expert systems that can 'automatically mine and detect patterns and build predictive models' based on large datasets (Kitchin, 2014: 101). These 'big data' methods are increasingly used in the analysis of governmental and business data,

scientific analysis and the analysis of social and cultural trends, constituting a major development in the field of data science.

Data science methodologies infuse the approach of Pearson's CDDAAL. On the 'meet the team' page of the CDDAAL website, for example, its director John Behrens is described as 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'. Staff are listed as 'research scientists' with expertise in data mining, computer science, algorithm design, intelligent systems, HCI, data analytics tools and methods, and interactive data visualization. In a methodological report for the CDDAAL, Behrens (2013) claims that educational research is increasingly under pressure to adopt new computational and data science methods that enable data manipulation and data visualization, including the mobilization of 'big data' to enable continuous tracking and monitoring of streaming data, rather than the collection of data through discrete temporal assessment events; the move towards 'population analytics' techniques that can handle enormous, scalable samples of many millions of records of research data; the use of 'educational data mining' to extract patterns from it; and the use of statistical models for combining results from different datasets and to integrate new and existing data and information.

In another CDDAAL publication, data science is a positioned as a 'transformative' methodology:

> Once much of 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. (DiCerbo and Behrens, 2014)

The authors argue that combining 'learning science' with data science methods will enable researchers to 'capture stream or trace data from learners' interactions' with learning materials; enable computer analysis to detect 'new patterns that may provide evidence about learning'; provide immediate feedback about performance on specific activities; construct data-based profiles and 'better models of learners' knowledge, skills and attributes'; 'tune' those models through continuously updated streams of data to ensure the inferences drawn from them are accurate and valid; 'to take a learner's profile of knowledge, skills and attributes and determine the best subsequent activity'; and, finally, 'to more clearly understand the micro-patterns of teaching and learning by individuals and groups'.

The vision pursued in the report is explicitly modelled on the idea of tracing individuals' 'activity streams' and is derived from social networking sites such as Facebook, where the activity stream is an integral part of the user interaction with the system—a constant trace of the user's production of content, status updates, comments, and so on. DiCerbo and Behrens (2014) expand the notion of the activity stream to suggest that the 'power of these streams lies in their ability to record change as it occurs', and that for the purposes of education 'they have the potential to indicate changes in learning, motivation and other characteristics of interest as they happen'.

Pattern recognition methods

As the above examples indicate, the CDDAAL aims to mobilize techniques of social network analysis to mine students' data for patterns, based on the understanding that, 'faced with a very large number of potential variables, computers are able to perform pattern identification tasks that are beyond the scope of human abilities … not only to collect information but also detect patterns

within it' (DiCerbo and Behrens, 2014). The computational method of pattern recognition operates by taking log files of a user's activity and then subjecting it to detailed analysis using various measures; data captured from a single individual's log file can then be synthesized with other users' log files to see if they can be combined into generalizable indicators of aspects of learning. To do this, the report details how pattern recognition analysis can be used to trace and match patterns in learners' activities:

> Learner interactions with activities generate data that can be analysed for patterns. … Performance in individual activities can often provide immediate feedback … based on local pattern recognition, while performance over several activities can lead to profile updates, which can facilitate inferences about general performance. (DiCerbo and Behrens, 2014)

The methodological development of pattern recognition techniques is a major strand of the CDDAAL's R&D programme. CDDAAL researchers are even engaged in detecting patterns from young people's activities outside of formal education, in online videogaming environments and social networking sites. As learners interact with systems and with other people, 'software records' every aspect of their activity so that as learners interact in digital environments, in formal and informal contexts, 'actionable data can be drawn from both':

> 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. (DiCerbo and Behrens, 2014)

The promise of pattern recognition methods promoted by Pearson is therefore not simply of better tracking of learners, but also the generation of new generalizable theories and models of cognitive development and learner progression. Those insights can then be made actionable as new software-based pedagogic products; Pearson is, of course, well positioned as an educational publisher to codify these insights in its own software applications for schools.

The logics of pattern recognition mobilized by the CDDAAL owe much to social media analytics from the commercial domain, and to the specific methods developed to detect, classify and extract associations and patterns from large datasets (Kitchin, 2014). Methods including cluster analysis, natural language processing, Bayesian networks, artificial neural networks and statistical analysis can then be used to find relationships between data objects, identify trends and curves, and make predictions about certain attributes on the basis of other attributes. CDDAAL researchers explicitly mobilize such pattern recognition methods to reveal the hidden patterns of learning and build generalizable models of cognitive development. For example, in another CDDAAL paper on the methodological challenges of analysing educational big data, Behrens (2013: 18) provides an upbeat assessment of how insights extracted from the generation of huge quantities of educational data will challenge current theoretical frameworks for making sense of it, 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'.

Fundamentally, the activities of the CDDAAL are premised on the big data epistemology that pattern recognition methods and techniques can reveal meaningful connections, associations, relationships, effects and correlations about human behaviours without the need for prior hypotheses, theoretical frameworks or further experimentation. This assumes that 'through the application of agnostic data analytics the data can speak for themselves free of human bias or framing, and that any patterns and relationships within big data are inherently meaningful and truthful' (Kitchin, 2014: 132). Yet, data do not exist naturally as a 'raw' or truthful representation of an underlying

reality; they have to be brought into being through social, methodological and technical practices, and are constantly shaped as they move between human actors, software platforms and institutional structures and settings, all framed by social, political and economic contexts (Bowker, 2005). Based on the epistemological assumption that pattern recognition software can reveal truthful models of human action, the CDDAAL aims to develop computational theories of learning itself as a means towards crafting better pedagogic techniques for governing learners.

Visual methods

Much of the research undertaken by CDDAAL researchers is highly technical in nature, traversing learning science and data science methodologies for mapping and modelling the generalizable patterns of learning processes and cognitive development. In order to make the insights it has extracted from these patterns in the data persuasive and acceptable to wider publics of policymakers, practitioners and even parents, Pearson has drawn significantly on data visualization methods to 'effectively reveal and communicate the structure, patterns and trends of variables and their interconnections' (Kitchin, 2014: 106). With the massive growth of digital data in education detailed by Pearson, visualization is employed to make visible and comprehensible complex datasets that would otherwise be difficult to conceptualize, and to reveal patterns, structures and interconnections that might otherwise remain hidden. Michael Barber has described how the Learning Curve supports 'evidence-based policy' through data visualization 'to make it easy for people … to use quickly without undermining the integrity of the data' (Barber with Ozga, 2014: 77).

For example, the Learning Curve mobilizes visual methods to reveal the patterns and associations between educational input and output indicators on a global scale. It features a suite of dynamic 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 and graphical time-series trend tools. It also permits the user to generate country infographic profiles that visually compare multiple education input indicators (such as public educational expenditure, pupil:teacher ratio, educational 'life expectancy') with education output indicators (PISA scores, graduation rates, labour market productivity), as well as socio-economic indicators (gross domestic product (GDP) and crime statistics).

Moreover, the Learning Curve is used as a form of visual argumentation. Through the application of visual analytics algorithms, it allows the user to manipulate the images in order to reveal patterns and associations, to conduct comparisons by altering variables, and to build visual models and explanations. The logic of country comparison underpinning the Learning Curve at least partly depends on the visual semiotics of the graphic presentation of patterns in the data. Visualization acts as a way of simplifying and reducing the complexity of the interaction of variables to graphical and diagrammatic form; it is an advanced semiotic technique of commensuration, whereby diverse quantities and qualities of educational data are transformed and standardized into a common visual metric. The methodological notes on the Learning Curve website are carefully worded to detail the data quality issues involved in aggregating its 60 different datasets; yet, its heatmaps, time-series tools and league tables smooth over these numerical problems to provide a glossy plane of graphical commensuration through which comparisons can be made and to which evaluations might be attached.

As this would suggest, the visualization of data is no neutral or objective accomplishment. Visual methods give the numbers meaning; they translate numerical measurements into curves and trends; they make the data amenable to being inserted into presentations and arguments that might

be used to produce conviction in others. In other words, data visualization gives numbers additional pliability to be shaped and configured as powerful and persuasive presentations. A data visualization is assembled as it circulates around a network of offices and computer screens, as it is worked on by designers, visualizers, project managers, programmers and data analysts, and as it moves between software programmes and hardware devices, 'through which data are constantly mobile, shifting and proliferating, moving between different actors and media, ported and patched, altered and designed, collaged and commented on' (Rose et al., 2014: 401). The human eyes and hands, as well as software platforms and algorithms, involved in its display shape the interpretations data visualization makes possible and the possible meanings that might be extracted from it. Visualization is thus also socially productive, in that it directs attention to correlations between data variables and objects that might then be made actionable as insights for decision-making.

The Learning Curve ultimately visualizes a virtual reference space against which all education systems might measure and monitor themselves; it constitutes a virtual comparator and a global benchmark for educational evaluation, judgement and action. It is through such visual techniques that Pearson seeks to attract various publics to the insights it has extracted from patterns of learning processes in its data, and to secure consensus that the models it has constructed from the data represent learning as it really is rather than as abstracted theories constructed from pre-existing disciplinary frameworks such as those associated with the social sciences.

Human–computer interaction methods

While visual methods of graphical data presentation enhance the plasticity of numbers, there is also considerable flexibility in its interpretation by users. The possible 'interpretive flexibility' available, however, is counterbalanced by the ways that data visualization itself 'configures the user' (Woolgar, 1991). Oudshoorn and Pinch (2003: 10–11) have detailed not only how 'designers inscribe their views of users and use in technological objects', thus constraining use of those technologies in particular prescribed ways, but also highlight how users might 'underwrite or reject and renegotiate the prescriptions'.

Reflecting the tension between the configuration of the user by the designer and the users' reciprocal reconfiguration of the designed object through resistant or unintended usages, the Learning Curve does not merely present ready-made visualizations, but also provides interactive tools to enable the user to conduct his or her own visual analyses through tweaking variables, selecting data sources and adjusting statistical weightings. Michael Barber has described it as a product of 'co-creation' that allows the public to 'play' with the data and 'connect the bits together' in a way that is more 'fun' than preformatted policy reports (Barber with Ozga, 2014: 84). The Learning Curve functions as an exemplar of a 'communication-based and information-based instrument' that privileges 'audience democracy' (Lascoumes and Gales, 2007: 13–14), whereby public authorities are obliged to provide citizens with rights of access to the information they hold and citizens are required to play a reciprocal role in its interpretation and dissemination.

However, the Learning Curve does not represent unconstrained audience participation. Within the design constraints of the Learning Curve, users are guided towards particular forms of analysis that privilege country comparison over other possible analyses, thus enabling and delimiting what users can do with the data and what can be said about it. Consonant with the comparative methods of the EIU that designed it, global comparison and forecasting—and the values and methodological preferences that underpin such approaches—are structured into the user interface through its league tables, heatmaps and time-series tools to shape interpretation, make visible particular educational realities and encourage particular kinds of responses. The design of the Learning Curve interface configures the research user as a comparative analyst and a data co-producer.

Approached in terms of methods, the Learning Curve is the product of emerging HCI methodologies. HCI methods have developed significantly in the context of big data, as:

> …information providers conduct a great deal of research trying to understand, and then operationalize, how humans habitually seek, engage with, and digest information. … [In HCI], the understanding of human psychology and perception is brought to bear on the design of algorithms and the ways in which their results should be represented. (Gillespie, 2014: 174)

Big data itself has become a valuable resource for HCI researchers and developers, who are able to utilize masses of data about users' information processing practices to inform the design of new software interfaces and functionality. Moreover, the aim of HCI in relation to commercial social media has been the optimization of the interaction in order to attract and seduce users to play a significant part in the continual production and circulation of informational and media content.

The logics of audience participation built in to the Learning Curve are not merely artefacts of a commitment to data transparency and democracy, as Michael Barber claims. They are the product of HCI insights about human information processing, perception and capacity to comprehend large data, twinned with the social media model of user interactivity and participatory content production. Through inviting user participation in the Learning Curve, it may even be possible for Pearson to track how users interact with the data—to monitor its efficacy, as its Efficacy Framework is intended—and then seek to optimize its tools to positively impact on future use. Yet, by including only internationally comparable data, the user's interactivity is already pre-figured by the Learning Curve, leading to a subtle but significant reinforcement of the methodological assumptions underpinning its interface and interaction design.

The kinds of user interaction experiences enabled by HCI methods have the capacity to configure the research user of the Learning Curve (the school leader, the practitioner, the policymaker, the researcher), yet the methodological advances in the HCI field underpinning the presentation and circulation of educational data, or its productivity to configure the user of the data, remain under-examined. This is a critical omission, as Pearson is rapidly developing the capacity to visualize its datasets and to amplify the public accessibility of the models of learning it is developing through its research at the CDDAAL and enforcing through the Learning Curve. As a consequence, the data-based models and classifications of learning and cognitive skills it is extracting from patterns in learning data have the potential to become more widely accepted as the reality of learning, and therefore to configure the practices of teachers and the decision-making of policymakers as its users.

Machine learning methods

The sixth part of Pearson's methodological complex is its capacity for prediction and pre-emption. In 2014 Pearson published a report on using 'intelligent software and a range of devices that facilitate unobtrusive classroom data collection in real time' (Hill and Barber, 2014). Its authors promote 'the application of data analytics and the adoption of new metrics to generate deeper insights into and richer information on learning and teaching', and to provide 'ongoing feedback to personalise instruction and improve learning and teaching' (Hill and Barber, 2014). Such systems, they argue, could instantiate a revolution in education policy, shifting the focus from the governance of education through the institution of the school to 'the student as the focus of educational policy and concerted attention to personalising learning' (Hill and Barber, 2014). Here, Pearson's ambitions are most starkly realized: it aims to make its emerging insights about learning, derived from patterns in masses of learners' data and translated into generalizable models of learning, into the

central focus of a predictive mode of educational practice and policy driven by machine-based intelligence. Pearson has even supported R&D in the area of artificial intelligence in education, claiming that 'artificial intelligence is increasingly present in tools such as adaptive curricula, online personalised tutors, and teachable agents' (Pearson College London, 2015).

The technical developments underpinning such an anticipatory approach are premised on currently emerging idea from technical R&D in 'learning analytics'. 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/)

Based on artificially intelligent machine learning algorithms, learning analytics software platforms like Knewton are designed to enable individual students to be tracked through their digital data traces in real time and to provide automated predictions of future progress (Siemens, 2013). Here, machine learning algorithms and the predictive analytics and prescriptive analytics they enact, are significant. Through machine learning techniques, 'programmers construct models that predict what people will do' by 'transforming data on events, actions, behaviours, beliefs and desires' into probabilistic predictions of the future that then can be used to decide on action to be taken in the present (Mackenzie, 2013: 399). Predictive learning analytics are one material instantiation of machine learning. Prescriptive analytics can then be mobilized as 'recommender systems' for personalized pedagogic intervention.

Consonant with the social life of the methods approach, it is important to acknowledge that learning analytics is itself a field of methodological inquiry, part of an 'emerging field of Educational Data Science' (Piety et al., 2014). The predictive models generated by the machine learning algorithms of learning analytics are therefore the product of complex social, technical and transdisciplinary practices and are embedded in the methodological commitments, assumptions, values and styles of thinking of their designers. As Piety et al. (2014) acknowledge in relation to educational data science, its 'architectures can encode various theories of learning that manifest themselves in the data the tools provide'.

Pearson's director of the CDDAAL, John Behrens, is a key voice in the field of educational data science (Piety et al., 2013). Shaped by the methodological commitments and constraints of the educational data science field, the predictive learning analytics techniques being developed through Pearson's partnership with Knewton anticipate a form of future-tense educational management through machine learning methods. These analytics capacities not only complement existing large-scale database techniques of governance conducted at discrete temporal intervals through large-scale testing, but also accelerate the timescales of governing by numbers. They make the collection of enumerable educational data, its processes of calculation and its consequences into a real-time and recursive process operationalized up close from within the classroom and regulated at a distance by new centres of statistical calculation, data analytics, pattern recognition, interactive visualization and prediction. Data are therefore being used 'to govern by activating the capacities of the individual' (Ozga et al., 2011: 88), a strategy accomplished through digital methods that capture, process and display highly granular detail on individual learners, their performances and their predicted progress in real time, and that then prescribe or recommend pedagogic interventions directly, rather than merely capturing snapshots of national systems at discrete temporal intervals.

As an 'open, living database', as Barber has described it, future iterations of the Learning Curve might also feature the kind of fine-grained individualized data on learners' cognitive skills that is becoming available through developments enacted by the CDDAAL too, enabling users to interact with the data aggregated from individuals' activity streams as well as conducting comparisons with time-series data from large-scale assessments. This would be consistent with Pearson's ambitions to shift the policy focus from large-scale testing to individuals' learning, and with its analytics capacities to generate generalizable models of cognitive skills development. The Learning Curve already commensurates diverse data sources to tabulate and visualize a common cognitive skills metric. If governing by numbers has been concerned with processes of commensuration and comparison across geographical regions, the digital governing methods of Pearson's analytics-based approach amplify its focus on the comparison of individual learners' cognitive skills within a global database that itself contains a generalized model of cognitive skills derived from massive populations of learner data.

Within this approach, then, Pearson is actively intervening in the classifications and categories by which learning is known, conceptualized and acted upon, enabling individual learners to be compared against algorithmic norms and globally standardized classifications of learning progressions. Its claim to be filling the gap between data-based results and the theory base to integrate them, leading to the production of new generalizable models of learning processes and progressions, means that its theoretical understandings and models of learning might then be transcoded into the pedagogic resources that Pearson itself produces and promotes to schools, particularly its personalized and adaptive learning applications. It is seeking to mobilize machine learning methods, as a form of artificial intelligence, to accomplish this task.

Governing methods

Methods socialize the objects of analysis. The methodological maximalism enacted by Pearson reveals the extent to which remediated digital methods derived from overlapping disciplinary traditions and epistemological perspectives are combining as a set of operational practices for the governing of education. By hybridizing the methodologies of data science and learning science (and their parent disciplines of CompSci and psy), Pearson has not only socialized these methods in the sense of normalizing them in educational inquiry, but also as a means towards wrapping new social norms, understandings and interpretations around education itself. The central contribution of this article has been to explore the consequences of Pearson's digital methods in terms of how its generation of new models and classifications of learning might then loop back into the pedagogic machinery of the classroom by being codified in software products. Four key points emerge from this analysis.

Firstly, Pearson's methodological complex constitutes a significant set of data practices for the production and performativity of an educational data/knowledge infrastructure, and influences the governing knowledge it produces. Through the work of both the Learning Curve and the CDDAAL, it enacts the enumeration of national and individual performance; the analysis and presentation of data; the embedding of data into practices, including not only policymaking but also pedagogic routines; the production of a virtual reference space that acts as a comparator model for other educational spaces; and the production of new practices, such as participatory data visualization and prediction. This infrastructure of data-based knowledge production is being constructed gradually, method by method, as well as line by line, out of software code and algorithms. The coding practices of programmers, software developers, algorithm designers and other technical experts are combined with data scientific research methods expertise in the development of Pearson's methodological complex and the data infrastructure it enacts.

Secondly, as an increasingly prominent actor in the production of a global educational data infrastructure, Pearson is positioning itself as an institutionalized governing expert, peopled by algorithmists with access to the methodological complex of software, algorithms, analytics and visualization tools required to analyse and make sense of the growing mass of data becoming available as education is digitized. Its remediated methods allow it to see, know and interpret aspects of education in ways that displace the authoritative knowledge of the educational psychologist, sociologist, historian or philosopher to the knowledge produced by the data scientist or even by automated machine intelligence. This reflects social scientists' anxieties about their authority to record and report on social phenomena at a time when big data methods have been elevated to methodological supremacy in commercial, cultural and political contexts, and social research has been redistributed to data science laboratories. Pearson is making data science methods into a new mode of expertise for knowing and intervening in learning, and is even seeking to transform the bureaucratic organs of official policymaking by emphasizing the personalization of learning at the individual level—directly through embedding adaptive software products in the pedagogic routines of the classroom—over incremental improvement to education systems.

Thirdly, through its redistribution and remediation of methods, Pearson is making educational data available to the eyes, hands and minds of educational policymakers and practitioners worldwide. Underpinning its approach is a data scientific commitment to data as a theory-free window on to educational realities, twinned with the epistemological assumption of the learning science field that education can be calculated objectively and visualized as existing facts about learning processes and cognitive development. However, its methods also subtly direct the gaze and guide the fingers to make sense of those data in particular ways, not just by simplifying the complexity of the data but by amplifying the perceptibility of certain features and reducing others to construct a new 'virtual world of educational data' (Lawn, 2013). Within this virtual world of data, learning is being reconfigured by Pearson in terms of commensurable data, patterns, predictions and visualizations that are themselves shaped by the social contexts in which they are produced. Its new data-derived models of learning and cognitive development have the potential to shape how pedagogic practitioners and policymakers understand what learning is and how to activate it through specific pedagogic resources, approaches and applications—all of which Pearson is itself positioned to provide as the world's largest publisher of educational resources and technologies:

> Pearson is involved both in seeking to influence the education policy environment, the way that policy 'solutions' are conceived, and, at the same time, creating new market niches that its constantly adapting and transforming business can then address and respond to with new 'products'. (Ball and Junemann, 2015: 7)

Informed both by the methods of data science and learning science, its products provide the means for the personalization of learning but 'at the same time can demonstrate impact in the form of measurable (learning) outcomes' (Ball and Junemann, 2015: 31).

Fourthly, however, these developments need to be understood not just in terms of commercial business models, but also as technical and methodological accomplishments that are both socially enacted and socially productive, and that are redefining how learning is understood and how learners are to be made amenable to pedagogic intervention through the technologies of schooling. The consequence is that Pearson is poised to exert a kind of 'looping effect' (Hacking, 2007) on learners' subjectivities, where the data-derived model acts to shape and 'make up' the people that it purports to measure and represent. In other words, the combined learning science and data science methods of Pearson could become highly consequential to the formation of new models of learning, and thereby to 'making up' students as new 'kinds of people' who are understood in

terms of the data and encouraged through the pedagogic apparatus of the adaptive classroom to relate to their own learning in novel ways. These developments amount to the production of a new data/knowledge infrastructure in education: the generation of classifications and categories of learning, measured and monitored in terms of the models produced by learning and data science practitioners, that might then 'touch people' (Bowker and Star, 1999: 314) by looping back into the classroom as pedagogic software applications. Methodologically, it is a matter of transcoding classifications of learning into the lines of code that constitute Pearson's e-learning software products—software code that then activates students' capacities in accordance with the codes of conduct contained in the classification.

Conclusion

Pearson plc has become an important policy actor with a 'network of interests and objectives' that stretch 'beyond corporate boundaries and into spaces of policy, academic research and philanthropy', through which it is strengthening its role 'across all aspects of the education policy cycle, from agenda setting, through policy production and implementation to evaluation' (Hogan et al., 2015: 62). Its high-profile personnel, such as Michael Barber, give Pearson policy credibility and leverage, whilst its presence in the field of educational data science through John Behrens positions it at the forefront of an emerging academic field of methodological inquiry and discovery. This article has surveyed key components of the methodological complex Pearson is inserting into the policy cycle—as well as into pedagogic practice—and explored the social life of its methods to underline its capacity to produce the data, analyses and knowledge required for the soft governance of education. It is achieving this through employing a range of data-based digital methods to produce new knowledge about learners and their learning processes, leading towards the production of a knowledge infrastructure within which new understandings and models of learning can be circulated and codified into pedagogic software products and recommendations. Pearson is also committed to monitoring the efficacy of its products, thus producing a highly recursive feedback loop that includes modelling and classifying cognitive learning processes, developing new pedagogic products to activate these processes, and then testing their effectiveness in terms of outcomes.

These activities are embryonic of an emerging hybrid of computer science-based data practices and psychological learning sciences that Loveless and Williamson (2013: 13) term a 'CompPsy complex', which 'assembles and encodes a particular representation of learner subjectivity' into specific technologies that are designed to 'elicit, promote, facilitate and foster the capacities, capabilities and qualities of such a pedagogic subject'. Mobilizing such a CompPsy complex, Pearson's Learning Curve, the CDDAAL, and other products such as its Efficacy Framework are key resources—each underpinned by very particular methodological techniques—for operationalizing the data infrastructure within which new and powerful classifications of learning derived from the learning sciences and computed through data will be produced and circulated as a kind of governing knowledge.

Pearson's work in education demonstrates how methods themselves are *socially produced*: they have past lives in disciplinary traditions; they are made by human hands with particular interests, guided by epistemological assumptions; they are selective, partial and always framed by the cultural, political and economic contexts in which they are deployed. Digital methods have become socialized as expert ways of knowing, seeing and evaluating social phenomena. As the social product of human endeavours, methods are also *socially productive*: they frame a problem to be seen in a particular way; they visualize results to shape interpretation; they produce classifications; and they direct attention towards particular patterns and translate associations and connections into

insights that might be acted upon through particular forms of social action. The social production and social productivity of new software-mediated digital methods now being deployed from the CompPsy laboratories of methodological experts such as Pearson require detailed interrogation as they exert social, political and material effects in digital education governance. Pearson's methodological complex is integral to its construction of new models and classifications of learning—a new knowledge infrastructure for knowing and acting in educational institutions—and represent the displacement of social scientific ways of knowing and intervening in the learning that takes place in schools to emerging data scientific modes of collecting, calculating and classifying learning.

Declaration of conflicting interest

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Ball S (2012) *Global Education Inc. New Policy Networks and the Neoliberal Imaginary*. London: Routledge.

Ball S and Junemann C (2015) Pearson and PALF: The mutating giant. Brussels: Education International. Available at: http://download.ei-ie.org/Docs/WebDepot/ei_palf_publication_web_pages_01072015.pdf (accessed 24 September 2015).

Barber M with Ozga J (2014) Data work: Michael Barber in conversation with Jenny Ozga. In: Fenwick T, Mangez E and Ozga J (eds) *Governing Knowledge: Comparison, Knowledge-Based Technologies and Expertise in the Regulation of Education*. London: Routledge, pp.75–85.

Behrens J (2013) Harnessing the currents of the digital ocean. In: *the annual meeting of the American Educational Research Association*, San Francisco, CA, April 2013.

Bowker G (2005) *Memory Practices in the Sciences*. London: MIT Press.

Bowker GC and Star SL (1999) *Sorting Things Out: Classification and its Consequences*. London: MIT Press.

Burrows R and Savage M (2014) After the crisis? Big Data and the methodological challenges of empirical sociology. *Big Data & Society* 1(1): 1–6.

Davies W (2015) The return of social government: From 'socialist calculation' to 'social analytics'. *European Journal of Social Theory*. Epub ahead of print 15 April 2015. DOI: http://dx.doi.org/10.1177/1368431015578044.

Decuypere M, Ceulemens C and Simons M (2014) Schools in the making: mapping digital spaces of evidence. *Journal of Education Policy* 29(5): 617–639.

DiCerbo KE and Behrens JT (2014) *Impacts of the Digital Ocean*. Austin, TX: Pearson.

Edwards P, Jackson S, Chalmers M, et al. (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor, MI: Deep Blue.

Edwards R and Fenwich T (this issue) Exploring the role of smart technologies on professional responsibilities and education. *European Educational Research Journal*.

Fenwick T, Mangez E and Ozga J (2014) *Governing Knowledge: Comparison, Knowledge-Based Technologies and Expertise in the Regulation of Education*. London: Routledge.

Gillespie T (2014) The relevance of algorithms. In: Gillespie T, Boczkowski PJ and Foot KA (eds) *Media Technologies: Essays on Communication, Materiality and Society*. London: MIT Press, pp.167–193.

Gorur R (2013) My school, my market. *Discourse: Studies in the Cultural Politics of Education* 34(2): 214–230.

Grek S (2009) Governing by numbers: the PISA 'effect' in Europe. *Journal of Education Policy* 24(1): 23–37.

Hacking I (2007) Kinds of people: Moving targets. *Proceedings of the British Academy* 151: 285–318.

Hardy I (2015) A logic of enumeration: the nature and effects of national literacy and numeracy testing in Australia. *Journal of Education Policy* 30(3): 335–362.

Hill P and Barber M (2014) *Preparing for a Renaissance in Assessment*. London: Pearson.

Hogan A, Sellar S and Lingard B (2015) Network restructuring of global edu-business: The case of Pearson's efficacy framework. In: Au W and Ferrare JJ (eds) *Mapping Corporate Education Reform: Power and Policy Networks in the Neoliberal State*. London: Routledge, pp.43–64.

Kitchin R (2014) *The Data Revolution: Big DATA, OPEN DATA, DATA INFRASTRUCTURES and their Consequences*. London: Sage.

Lascoumes P and le Gales P (2007) Introduction: Understanding public policy through its instruments—from the nature of instruments to the sociology of public policy instrumentation. *Governance* 20(1): 1–21.

Law J, Ruppert E and Savage M (2011) The double social life of methods. CRESC Working Paper No. 95, Open University.

Lawn M (2013) The rise of data in education. In: Lawn M (ed.) *The Rise of Data in Education Systems: Collection, Visualization and Use*. Oxford: Symposium, pp.7–25.

Lawn M and Grek S (2012) *Europeanizing Education: Governing a New Policy Space*. Oxford: Symposium.

Loveless A and Williamson B (2013) *Learning Identities in a Digital Age: Rethinking Creativity, Education and Technology*. London: Routledge.

Lupton D (2015) *Digital Sociology*. London: Routledge.

Mackenzie A (2013) Programming subjects in the regime of anticipation: Software studies and subjectivity. *Subjectivity* 6(4): 391–405.

Marres N (2012) The redistribution of methods: On intervention in digital social research, broadly conceived. *Sociological Review* 60(S1): 139–165.

Mayer-Schonberger V and Cukier K (2013) *Big Data: A Revolution that will change how we Live, Work and Think*. London: John Murray.

Moos L (2009) Hard and soft governance: The journey from transnational agencies to school leadership. *European Educational Research Journal* 8(3): 397–406.

Oudshoorn N and Pinch T (2003) How users and non-users matter. In: Oudshoorn N and Pinch T (eds). *How Users Matter: The Co-construction of Users and Technology*. London: MIT Press, pp.1–25.

Ozga J, Segerholm C and Simola H (2011) The governance turn. In: Ozga J, Dahler-Larsen P, Segerholm C et al. (eds). *Fabricating Quality in Education: Data and Governance in Europe*. London: Rouledge, pp.85–95.

Pearson (2014) Pearson to develop PISA 2018 Student Assessment 21st Century Frameworks for OECD. Pearson. com, 10 December 2014. Available at: https://www.pearson.com/news/announcements/2014/december/pearson-to-develop-pisa-2018-student-assessment-21st-century-fra.html(accessed 24 September 2015).

Pearson College London (2015) Intelligence Unleashed: How smarter digital tools can improve learning. Pearson College London, 7 September 2015. Available at: https://www.pearsoncollegelondon.ac.uk/subjects/business-school/news/2015/09/intelligence-unleashed.html (accessed 24 September 2015).

Piattoeva N (2015) Elastic numbers: national examinations data as a technology of government. *Journal of Education Policy* 30(3): 316–334.

Piety PJ, Behrens J and Pea R (2013) Educational data sciences and the need for interpretive skills. American Educational Research Association, 27 April–1 May 2013, San Francisco. Available at: http://www.slide-share.net/PhilipPiety/education-data-sciences-and-interpretive-skills (accessed 24 September 2015).

Piety PJ, Hickey DT and Bishop MJ (2014) Educational data sciences—framing emergent practices for analytics of learning, organizations and systems. In: *LAK '14*, 24–28 March 2014, Indianapolis, IN. Available at: http://edinfoconnections.com/wp-content/uploads/2014/01/Educational-Data-Sciences-Feb-9.pdf.

Popkewitz TS (2012) Numbers in grids of intelligibility: Making sense of how educational truth is told. In: Lauder H, Young M, Daniels H, et al. (eds) *Educating for the Knowledge Economy? Critical Perspectives*. Abingdon: Routledge, pp.169–191.

Rogers R (2013) *Digital Methods*. London: MIT Press.

Rose G, Degen M and Melhuish C (2014) Networks, interfaces and computer-generated images: learning from digital visualizations of urban regeneration projects. *Environment & Planning D: Society & Space* 32(3): 386–403.

Rose N (1999a) *Governing the Soul: The Shaping of the Private Self*. 2nd ed. London: Free Association Books.

Rose N (1999b) *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press.

Ruppert E, Law J and Savage M (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30(4): 22–46.

Ruppert E, et al. (2015) Socializing big data: from concept to practice. CRESC Working Paper no. 138. Available at: http://www.cresc.ac.uk/medialibrary/workingpapers/wp138.pdf (accessed 24 September 2015).

Savage M (2013) The 'social life of methods': A critical introduction. *Theory, Culture & Society* 30(4): 3–21.

Sellar S (2014) Data infrastructure: A review of expanding accountability systems and large-scale assessments in education. *Discourse: Studies in the Cultural Politics of Education* 36(5): 765–777.

Sellar S (2015) A feel for numbers: Affect, data and education policy. *Critical Studies in Education* 56(1): 131–146.

Siemens G (2013) Learning analytics: The emergence of a discipline. *American Behavioral Scientist* 57(10): 1380–1400.

Williamson B (2015) Digital education governance: data visualization, predictive analytics and 'real-time' policy instruments. *Journal of Education Policy*. Epub ahead of print 15 April 2015. DOI: http://dx.doi.org/10.1080/02680939.2015.1035758.

Woolgar S (1991) Configuring the user: The case of usability trials. In: Law J (ed.) *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, pp.57–99.

Author biography

Ben Williamson is a lecturer in the School of Education at the University of Stirling. His research focuses on digital data, software code and algorithms in the technologies and practices of education governance, and on the influence of think tanks and policy innovation labs in education policy. His recent articles have been published in the *Journal of Education Policy, Critical Policy Studies*, and the *Journal of Educational Administration and History*.