Education as a Domain of Natural Data Extraction: Analysing Corporate Discourse about Educational Tracking

Jun Yu

London School of Economics and Political Science

Nick Couldry

London School of Economics and Political Science

Abstract

Digital platforms and learning analytics are becoming increasingly widespread in the education sector: commercial corporations argue their benefits for teaching and learning, thereby endorsing the continuous automated collection and processing of student data for measurement, assessment, management, and identity formation. Largely missing in these discourses, however, are the potential costs of datafication for pupils' and teachers' agency and the meaning of education itself. This article explores the general discursive framing by which these surveillant practices in education have come to seem natural. Through a study of commercial suppliers of educational platforms, we show how the prevailing vision of datafication in their discourses categorises software systems, not teachers, as central to education, reimagining space, time, and agency within educational processes around the organisation of data systems and the demands of commercial data production. Not only does this legitimate the new connective environment of dataveillance (that is, surveillance through data processing), but it also naturalises a wider *normative* environment in which teachers and students are assigned new roles and responsibilities. In the process, the panoptic possibilities of ubiquitous commercial access to personal educational data are presented as part of a virtuous circle of knowledge production and even training for good citizenship. This broader rethinking of education *through* surveillance must itself be critiqued.

Keywords: big data, datafication, data collection, learning analytics, surveillance, dataveillance

Introduction

Datafication – the increasingly continuous collection and processing of data from everyday life-streams and transactions – is transforming social life in all its aspects, from consumption to casual interaction. Discourses around 'Big Data' are of particular importance to this transformation (Mayer-Schönberger & Cukier, 2013; Kitchin, 2014). One particular area is education, the specialised domain where subjects are formed as citizens, workers and moral beings. A number of scholars (Breiter & Jarke, 2016; Selwyn, 2015a, 2016a; Taylor, 2017; Taylor & Rooney, 2017; Williamson, 2015) have raised concerns about how digital educational platforms are collecting data from and around the classroom, raising questions about the implications of such platforms for children's human rights (Lupton & Williamson, 2017). While such extended forms of data collection *might* have beneficial consequences if consensually implemented, what if the *discourse* of such platforms serves to naturalise surveillance and data extraction in the educational process? Drawing on a longer literature on category naturalisation's role in social order, we uncover this naturalisation at work in the language of eight leading educational platforms and the wider business discourse about data on which those platforms draw.[1] This article contributes to the critical sociology of data practices, while drawing on wider literatures in science and technology studies, critical data science and educational studies; through our close attention to the language of naturalisation, we also contribute to a broader understanding of the rhetoric of digital societies.

This article follows a much longer critical debate in educational studies about surveillance in schools, especially for security and behaviour control (Andrejevic & Selwyn, 2019; Casella, 2003; Deakin, Taylor, & Kupchik, 2018; Taylor, 2018). The impact, particularly in the USA, of the commercial security industry and a broader

culture of militarisation and neoliberal governance has been noted (Giroux, 2015; Lewis, 2003; Saltman, 2014), and finds echoes at points in our discussion, but falls short of installing surveillance *as* a technique of education. Meanwhile the introduction of surveillance techniques as tools for educational development via digital platforms has so far generated little public debate. Digital educational platforms run by commercial providers depend for their data processing on the continuous 'dataveillance' (Clarke 1988; Van Dijck 2014) of children. They give unprecedented agency to platform and software developers and data analysts in the educational process. Thereby the space of the classroom is being re-conceived, de-centring the teacher and students and shifting focus to software systems that offer continuous online performance monitoring, which the teacher and the educational system are required to support.

We examine this emerging, data-driven discourse about education, and show that it ignores the potential costs of datafication for children, teachers, and wider society. Promulgated by leading representatives of global business such as the World Economic Forum and global management consultancy firms, and applied by the corporations that sell digital platforms for education management, this discourse is reimagining education – its spaces and times – as the 'naturally' captured and managed domain of data systems (Agre, 1994). The agency of teacher and student in classroom interaction becomes displaced.

Our goal is not to critique the existence of surveillance in educational development (that has already been done), but to trace the evolution of its supporting discourses, and consider how the *intense embedding* of software platforms in everyday teaching might serve to naturalise practices of surveillance *still further*. Our argument proceeds, first, by considering surveillance's relation to the educational process. In the second section, we unpack the discourses of educational platform providers and trace their intersection

with broader business discourses about data collection. In the third section, we focus on how such corporate discourse rethinks the educational process, naturalising datafication and conflicting with underlying values of human autonomy.

The Expanding Role of Surveillance and Dataveillance in Education

Collecting data from children through close monitoring has long played a role in promoting their educational progress (Jenks, 2005), but, in some countries, specific surveillance processes within schooling have attracted criticism, including school and university league tables and globally standardised tests, such as the Programme for International Student Assessment (PISA), and the monitoring in UK schools of children's movements and activities through CCTV cameras and radio-frequency identification chips attached to school uniforms (Taylor, 2013). In the USA and UK, schools have become oriented towards 'making students visible and controllable' (Gilliom & Monahan, 2013, p. 74), and marked by high, if intermittent, surveillance of students (Williamson, 2017, p. 56). As such, the education sector offers a markedly different starting-point for discussion of surveillance from other sectors, such as health, that have long been characterised by sensitivity to the privacy of individual data.

In this article, our focus is North America and the UK, and a new and distinctive development: the introduction of something *close to* surveillance into basic processes of teaching, evaluating and monitoring student progress. Recalling the classic discussion of Roger Clarke (1988), we define 'surveillance' as 'the systematic investigation or monitoring of the acts or communications of one or more persons.' Through 'dataveillance' (Clarke's invented term), surveillance is achieved by the use of 'personal data systems' and the processing of the data they hold, rather than by human

beings or their tools watching or listening to other humans. New digital technologies that enable *continuous and automated* collection of personal data from and around the classroom involve the processing of data, as opposed to just watching what goes on in the classroom: this is dataveillance by Clarke's definition (Williamson, 2017). Contemporary dataveillance overcomes the limits of bodily surveillance in the classroom where the teacher's body could never achieve the total capture and recall required by panoptic surveillance (Gallagher, 2010). In a connected environment, large data systems *can* achieve total capture and recall. The goal of collecting and processing data *continuously* from students' educational performance now involves in-classroom apps and self-trackers such as ClassDojo and Sqord (Williamson, 2015), and 'learning analytics' that analyses students' performance to offer a fully 'personalised' education and learning environment (Luckin, Holmes, Griffiths, & Forcier, 2016).

This environment can offer individually tailored predictions of future student progress, and *pre-emptive* pedagogic intervention, that may sometimes prove beneficial.[2] But dataveillance at this level of intensity, potentially, reshapes what contemporary education *is*. Data collected from students are increasingly used to diagnose students and predict their future progress and risks (Williamson, 2015), affecting the distribution of resources (Breiter & Jarke, 2016). Students may be 'categorised' as either 'effective' or 'deviant' (Selwyn, 2014, p. 52) based not on judgement from teachers' lived experience but on the pattern detection of *data analytics*, potentially overriding 'the social factor . . . [that is] the mutual interaction of different minds with each other' in the classroom that John Dewey (1977, p. 261) saw at the heart of meaningful educational experience.

How has this *de facto* shift in educational authority and practice come to seem natural?

Project Design

In a popular business text, Mayer-Schönberger and Cukier (2014, p. 5) predicted that data will radically 'reshape learning' through datafication. More critical scholars fear the 'recursive state where data analysis begins to produce educational settings, as much as educational settings producing data' (Selwyn, 2015a, p. 72), turning schools into data platforms (or 'dataveillance school[s]' (Williamson, 2017)) linked to vast data collection programmes. To understand the discourse that might authorise this transformation, we must look more broadly at the *general framing* of data collection in contemporary societies.

We based our methodology on the conceptual tool of 'categories' (Bowker & Star, 1999), which refers to the everyday institutional process which 'decid[es] what will be visible and invisible within the [everyday] system' (Bowker & Star, 1999, p. 44). The naturalness of entities in the education sector emerges through how data and actors are categorised within accounts of good education along what Bowker and Star (1999) call a 'trajectory of naturalization' (p. 299). Naturalisation is more than mere routinisation of specific processes as its trajectory entails *'forgetting'* certain older aspects of education so that new categories (of data or education actor/action) can become visible *as natural*, and eventually unquestionable (Bowker & Star, 1999, Chapter 8; Chan, forthcoming).

To uncover the trajectories of naturalisation regarding data collection within the large sector of education, we conducted an analysis of public discourse of eight major publishing corporations, suppliers of learning analytics and social media platforms for children, and research institutions in education industry: Blackboard, IBM Watson

Education, Impero Education, Knewton, Macmillan Education, Microsoft's Educated Cities, Sqord, and Pearson Education.[3] The reports and websites studied here offer recipes for categorising educational practices on the ground. These corporations were selected to ensure a range of educational software providers active in North America and the UK (North America is the global leader in this area, whereas the UK is another country of fast development in digital education services, where, as researchers, we are based); to include a range of organisation scales (from IBM to Sqord); and to find companies whose public documents provided a range of discourses around their educational provision (e.g. mission statement, annual report, marketing language, blog posts, website text).

We analysed these documents after first sensitising ourselves through a review of the discourse on Big Data by leading representatives of global business, such as the World Economic Forum (WEF) and the Organisation for Economic Cooperation and Development (OECD), as well as reports for governments of a number of Western countries, undertaken as part of a larger project on discourses about Big Data and surveillance/dataveillance. That wider review sensitised us to certain trends in how dataveillance in society is, or is not, discussed. But in giving special attention to education as a domain – the domain where mature human subjects are formed, with potentially profound consequences for wider society and values – we sought to establish the extent to which those general trends were repeated, while also being open to new patterns of discourse distinctive to the education domain. To do so, we developed a coding framework that enabled us systematically to track both general and education-specific discourse relating to data collection and processing, whether or not it referred explicitly to issues of privacy and autonomy.

Through this analysis, we identified in education a dominant discourse around surveillance (dataveillance) that helps frame panoptic possibilities *positively*, constituting surveillance as an essential tool of education. By assuming that citizenship in the digital era means being countable along numerous dimensions at the 'demand' of external parties (Bowker, 2005, p. 30), it is suggested that continuous surveillance should extend outside the classroom and bridge pedagogic experience and wider life. While this discourse of connectedness may appear to exemplify the process that Basil Bernstein (2001) once identified positively as the 'total pedagogisation of society,' on closer inspection it merely installs *corporate* data actors as central to pedagogic performance and education.

Throughout our study, we were as concerned to look for patterns in what is *not* said, as in what is said, since structured absences within discourse can be significant.[4]

Educational Discourses about Data Collection

Let us first consider Big Data discourse from accredited representatives of global business, and then see its applications within the education sector.

Data as Natural Resource

The most fundamental move in today's dominant commercial discourse is to promote the idea that data and its growth are *natural*.[5] Data is categorised as a *raw material* with value. As such, data are 'the new 'oil' – a valuable resource of the 21st century' (WEF, 2011, p. 5), and a 'tradable asset . . . [which] must flow [freely] to creative value' (WEF, 2012a, p. 5). This 'natural resource' only has value if it is *used* well: 'data have

no intrinsic value; their value depends on the context of their use' (OECD, 2015, p. 197; italics added).

Categorising data as a naturally existing resource and emphasising its *use*-based value, relies on the fundamentally problematic idea of 'raw data' (Gitelman, 2013) that obscures the artificial processes of generating and collecting data from persons, and automatically reinforces the naturalness of appropriation and procession by corporations to create value. Metaphors that see data as mere 'exhaust' (UN, 2012) from everyday life underline this idea that data are not referable to personal ownership:

> in contrast to the concept of ownership of physical goods, where the owner typically has exclusive rights and control over the good – including for instance the freedom to destroy the good – this is not the case for intangibles such as data . . . The digital divide isn't about who owns data – it's about who can put that data to work (OECD, 2015, pp. 195–197).

This leads to a more controversial claim that it is only data use – not the underlying act of data collection – that creates concerns among citizens (White House, 2014, pp. xii–xiii). Potential concerns regarding data use are acknowledged (for example, by WEF: 'their use . . . can both generate great value and create significant harm, sometimes simultaneously' [WEF, 2013, p. 3]). However, the result is to bypass *separate* concerns about the collection of data, since data harvesting is assumed natural, a phenomenon of 'collective intelligence' (OECD, 2015, p. 352). Yet the assumption that data can 'speak for themselves,' free from human bias, positionality, or pre-determined framing, has been effectively deconstructed (boyd & Crawford, 2012; Van Dijck, 2014).[6]

Educational companies reproduce this view of data as natural resource:

> The proliferation of technology has created more data and at the same time has made it more accessible; educators *just* need the tools to put it to work to shape more personalized learning (IBM Watson Education; italics added).

A new type of data expertise is authorised in those institutions that sell digital platforms for education and use data as their raw material:

> It is this aggregated data set [from multiple sources] that forms the *raw material for our data scientists* (Blackboard; italics added).

These educational pronouncements extend the general discourse, while repeating its inattention to how data are generated and collected. Other educational players go further:

> as the daily activities of life become increasingly instrumented *without requiring additional effort*, we will move from performances completed primarily to collect data to performances completed for meaningful outcomes with data as a side effect (Pearson, 2014b, p. 15; italics added).

In the education sector, data is discussed in ways that fit with the bigger corporate vision of human life as a supposedly *natural* domain of data collection and processing.

The Naturally Connected Environment of Datafication

The next move is to naturalise the specific processes – of data*fication* – that generate benefits from this natural resource. Again let us start with general corporate discourse.

An important metaphor is that of a 'hyperconnected world' (WEF, 2012b), where modes of data collection are automated, and collected data goes beyond personal data:

> the exponential growth of mobile devices, big data, and social media are all drivers of this process of hyperconnectivity . . . This vision of our future hyperconnected world builds on the connectivity and functionality made possible by converged next-generation networks . . . Today, connected humans are already in the minority of Internet users (WEF, 2012b, pp. xi, 47).

This view of data collection as just 'automatic' facilitates an imaginary of data-driven *social* transformation that makes little, if any, reference to the price *people* might pay for their connectivity. What matters is the potential value growth of data pools that are 'reused to generate value. Data grow ever more connected and valuable with use' (WEF, 2012a, p. 7). This continuous loop of data processing and data aggregation is seen as 'a virtuous circle' (WEF, 2012b, p. 6), so authorising – indeed requiring – the free flow of data in space and time, and 'the fundamentally open and interconnected nature of information systems and networks' (OECD, 2015, p. 209).

When we turn to education platform providers, this general discourse is applied by imagining the classroom as a virtuous 'circle,' that is, an intensely connected place within a cycle of action-intervention-correction based on data resources previously unavailable to teachers:

> Learning Analytics predicts learner success with real-time data: . . . When an instructor sees that a learner's not taking part in online discussion forums, and is likely at risk for dropping a course, they can immediately intervene in a highly personalized way (Blackboard).

In effect, this assumes a dataveillant classroom, linked by systems of continuous data collection within a wider world of dataveillance: 'the means of data collection are increasingly embedded in the fabric of modern life' (Pearson, 2014b, p. 3). Such discourse links change in the corporatised classroom (Giroux, 2015) to wider changes in everyday social life and its massively enhanced infrastructure of data collection and processing:

> The app, IBM Watson Element for Educators, enables a new level of engagement for teachers by providing a holistic view of each student at their fingertips, including data on interests, accomplishments, academic performance, attendance, behaviors and learning activities (IBM, 2016).

> Easily reach your students, wherever they are (Blackboard).

Questions of privacy, while noted in passing, are secondary in this discourse: 'while privacy concerns one's *right* to keep information hidden, security may be an even greater concern, as it addresses one's *ability* to keep information hidden' (Pearson, 2014b, p. 23; italics in original). Against the background of this discourse about Big Data's *natural* role in the everyday world (across global business and the education

sector), let us explore how corporate discourse in education further naturalises practices of dataveillance in the classroom.

Reimagining Education Through Datafication

Here, our exclusive focus will be on the discourse of corporate educational providers. Data actors in the UK and US education industry are re-categorising themselves as the focal-point around which education *as a whole* should be organised, reimagining education around the needs of data systems and commercial data production. This discourse dismisses the educational model of the *pre-datafication* era as irresponsible and outdated, so further naturalising the new dataveillance-based model of education.

Reimagining Space and Time in Education

The corporate providers of educational platforms claim to offer a fully 'personalised' education for each student. Termed 'learning analytics' (Eynon, 2013), this education involves personalising learning environments through 'the [continuously adaptive] measurement, collection, analysis and reporting of data about learners and their contexts' (Siemens & Gasevic, 2012, p. 1). This new vision of learning is dependent on continuous monitoring and data processing:

> Adaptive learning works by assessing student performance and activity *in real time*. Then, using data and analytics, it personalizes content to reinforce concepts that target each student's particular strengths and weaknesses (Pearson; italics added).

As a result, learning analytics providers such as Knewton can claim to be 'absolutely data-driven.' This new model of datafied learning contrasts itself with a traditional chalk-and-board classroom:

> educators are building classrooms in the cloud. With *anywhere-anytime* access to learning materials, students and teachers move seamlessly from chalk-and-talk to social networks to online tools for shared social learning, presenting and collaborating, connecting the minds of education institutions in the Cloud (IBM Education, 2012a; italics added).

The key word here is 'anywhere-anytime,' a configuration based on the 'naturally' tracked space-time of data systems and installing IBM's centralised infrastructure (IBM, 2009).

In envisioning this reconfiguration of educational authority, explicit reliance is made to collected data's supposedly *natural* status:

> The collection of data and the corresponding gains of self-awareness, self-reflection, organisational evolution, and institutional insight are emerging from the bases of our *everyday natural activity* . . . It is this fundamental shift from data as the goal of our activity to data as a *side effect* of our activity that opens new doors to understanding and improving education (Pearson, 2014b, pp. 26–27; italics added).

Once human interactions and activities are seen as *naturally* fused with data-collection systems, optimising data outcomes follows naturally. In an environment of assumed dataveillance, students' own mentality is also supposedly transformed:

> Engage today's students on the devices they know and love, any time, any place. Our solutions are built with a mobile first mentality to offer students an engaging and flexible experience to maximize learning (Blackboard).

Through the steady unpacking of real-time monitoring and connectivity's implications, a new space and time for educational practice becomes not just imaginable, but practical.

Datafication and the Responsible Educator/Student

The result is to imagine a *normative* environment where both educators and students have new responsibilities. In this environment, Pearson claims for itself the role of 'stewardship' of others' learning (Pearson, 2014b, p. 1). This stewardship involves a constant attentiveness and receptivity to the *datafied* education process:

> the devices we interact with are typically designed to record these *fleeting* experiences [of digital life], creating a slowly rising ocean of digital data (Ibid.; italics added).

This new 'adaptive' learning is categorised as superior to the 'fleeting' signals of older forms of education that have proceeded for centuries *without* datafication. Metaphors such as 'factory-model' (Frase, 2014), 'digital desert' (Pearson, 2014b), or 'black-box' (Pearson, 2014c) present the historic *lack* of dataveillance techniques as random and

irresponsible, an irresponsibility 'resolved' by the fully 'personalised' education that datafication ensures:

> "adaptive [learning]" . . . is absolutely data-driven . . . You have to understand content and proficiency of students. And if you don't, you can build any kind of recommendation engine you want, but you're literally spitting out randomized answers, and that's completely irresponsible (Knewton).

> On the education side, we have tended to a factory model where all children receive the same content in the same manner, despite ample evidence of the value of personalized interventions aligned with learning style (Frase, 2014).

> For so long, much of what happened inside classrooms has remained hidden in a 'black box', making it difficult to pursue a deliberate and continuous approach to the improvement of learning and teaching (Pearson, 2014c, p. 56).

These three accounts differ in how explicitly they characterise the pre-datafication model of education as unethical, incompetent, or beyond improvement. Yet together they reinterpret the old education model within a historical teleology that culminates in a supposedly better model of education. Datafication is presented as education's inevitable *next* stage, so insulating it from ethical questioning and installing data companies as the natural *drivers* of educational change (Chan, forthcoming).

This discursive move appropriates to data companies unprecedented power within the educational process. Underlying it is a broader belief in the disruptive innovative power of new digital technologies as a supposedly neutral agent of improved learning

(Selwyn, 2016a), thereby *de*naturalising earlier educational environments, with their distinctive rhythms of and intervals between assessment:

> Today, students walk into classrooms each September as if they were just born. Teachers must learn everything about them from scratch. Knewton-powered apps change this, allowing each student to start courses "warm" by connecting his or her learning history to every app . . . student engagement can be strengthened if academic work is imbued with a sense of continuity (Knewton, 2014).

This double naturalisation/denaturalisation seeks to align forward-thinking teachers with the corporate agendas of data-driven capitalism.

Reimagining Pedagogic Agency in the Datafied Classroom

Pearson Education's mission is 'to help people make *measurable* progress in their lives through learning' (Pearson; italics added). Few would oppose entirely the necessity for some measurement in education, but what are the implications of this new datafied environment for the young human subjects tracked within it?

The implications, in particular, for privacy and autonomy are generally ignored in corporate discourses. Datafication is presented as enabling a rethinking of the skills that education teaches and how they align to wider economic objectives, as in IBM's vision:

> Historically it has been hard for businesses to forecast their skill requirements with enough precision and enough lead time for educational institutions to align to those needs. Without this visibility, it has been hard for parents, teachers or counsellors to

> give students useful advice on employable skills. However, analytic techniques now exist to help businesses include human resources and skills in their strategic planning, and to understand what the real skill characteristics will be (Frase, 2014).

Other actors, it is implied, must rethink the notion of 'learning' itself, with a focus on the 'personalisation' that embeds data collection in modern life:

> in an era when learning is distributed [i.e. a geographical separation of instructor and student for part or all of the learning experience], technology is allowing students to learn anytime, anywhere. These new entry points to higher education make possible more personalized measurements of learning than ever before (Blackboard, 2016).

While the discourse of 'personalisation' in education has been noted by others (Selwyn, 2016b), its implications for educational agency need further reflection.

Meanwhile, *system* knowledge on this view becomes as important to educational outcome as individual knowledge, whether of students or teachers:

> the information generated by learning systems will have value well beyond the individual learner: it will provide a source of *generalisable new knowledge*, paving the way for a 'design science' approach, in which the primary focus of educational research is on evidence-based strategies for improving learning and teaching (Pearson, 2014c, p. 56; italics added).

Of course, marketers of new data-driven educational tools are careful not to present them *explicitly* as supplanting the agency of teachers, as '[t]hese systems will play the

role of an assistant, which is complementary to and *not a substitute* for the art and craft of teaching' (IBM Education, 2016; italics added). However, datafication is being presented here as the builder of a new form of social *order* in educational institutions (Breiter & Jarke, 2016). The role and duties of teachers become to facilitate and implement data-based teaching driven by the new data platform:

> In next-generation learning systems, the teacher retains the key role in fostering the learning for each student, but the job itself changes. Learning systems of the future will free up teacher time currently spent on preparation, marking and record-keeping and allow a greater focus on the professional roles of diagnosis, personalised instruction, scaffolding deep learning, motivation, guidance and care (Pearson, 2014c, p. 56).

The primary emphasis shifts from teachers to the 'learning system' architecture. As Selwyn notes (2014, p. 52), in this data-based educational system, the knowledge about whether a student becomes an 'effective' or 'deviant' agent becomes sourced primarily through observation of data, relegating teachers to a supporting role in the process of system measurement.

The *connective classroom* environment can still be presented as enhancing teaching and learning benefits, but while also mystifying the role of technology:

> we need teachers to teach to individual students as much as they can; at the same time, we need **non-judgmental personalisation** that feeds into student's interests and aspirations. This does not lead to an opposition of the type relationship vs. algorithms, but might point instead to a fruitful collaboration between **human**

> **teachers and intelligent, data driven assistants** . . . *Technology isn't the focal point here, it's just a way to connect* ***one teacher*** *with* ***thirty students'*** *interests, hopes and dreams* (Gioga, 2016; italics added, bold in original).

Yet the idea that data processes – and their algorithmic categorisations – do not *already* 'judge' those whose data they collect and analyse has been comprehensively criticised also in the broader information science literature (Gillespie, 2014; O'Neil, 2016; Sandvig, Hamilton, Karahalios, & Langbort, 2016). The consequences for education more generally are beginning to be noticed (Lewis & Holloway, 2019).

Meanwhile, the corporations that collect and process educational data acquire an inflated agency. In a US Department of Education video, the CEO of Knewton, Jose Ferreira, called education 'the world's most data-mineable industry by far,' and claimed 'we have five orders of magnitude more data about you than Google has . . . we literally have more data about our students than any company has about anybody else about anything' (quoted in Hill, 2014). Let us turn to how this new corporate agency connects directly with a wider vision of surveillance.

'Real-time Visibility': A New Educational Value

Continuous datafication's role in the education of children has become so deeply naturalised that its panoptic possibilities are no longer even being disguised. Rather, they are being emphasised by corporations as the means to 'help teachers foster meaningful engagement in the classroom by providing a 360-degree view of every student . . . enabling them to deliver a personalised approach to learning' (IBM Watson Education)! Surveillance capacities become a *selling-point* of new systems: 'designed to

provide teachers with greater, hands-on control over the digital classroom . . . [via] real-time visibility of every user's device . . . in one single, central view' (Impero Education).

The already noted virtuous cycle of educational action-intervention-correction returns as the basis of judgement and action, but enhanced by the new 'visual' (in fact dataveillant) capacities of educational platforms:

> Our digital system gives you real-time visibility of every student's device in one clear, uncluttered, and intuitive central view. You can restrict or monitor access to websites, applications, and hardware (Impero Education).

A particularly interesting application of this general discourse of surveillance in education is the new educational discourse of 'good digital citizenship' (Impero Education, 2016). Here the link between continuous surveillance and the management of children becomes so direct that the former is not even defended, but rather *affirmed* as the tool towards a new *competence* for children within a new vision of educational *freedom*:

> Real-time monitoring is not about policing kids. Rather, it's about providing opportunities for mentorship, teaching and learning. Keyword detection, photo and video capture and logged incident reports provides educators and administrators with tools to mentor good digital citizenship . . . This allows students to be responsible, safe and good digital citizens – both in school and out in the world (Impero Education).

This new vision of *system*-based education assumes that school and childhood are an extreme risk environment which only datafication (with its own risks and costs ignored) can 'safeguard.'

It is here we find echoes of wider securitisation discourse in schools (Taylor, 2018). From this securitised starting-point, it is easy to justify continuous surveillance as a natural support to the educational process:

> Impero Education Pro has been specially developed to help schools keep students safe and protected in the online learning environment. That's why when your child attends an Impero school, you can be certain that they're in safe hands . . . Impero Education Pro also allows teaching staff to monitor students' online activity from their screen in real-time. A thumbnail view of all student screens, in one central view, allows potential risk (or other instances of misconduct) to be dealt with as and when they occur – just like any other behavioural issue. Our software also provides a complete log of all online activity (Impero Education).

The outcome is explicitly panoptic ('all student screens, in one view'), yet the widely-known risks of surveillant environments are here sublimated in a wider IT solution. Surveillance, far from having costs, is presented as the necessary *basis* of a new practice of 'digital citizenship':

> the idea [of digital monitoring] is to allow students the online freedom they need to grow, learn and survive in a digital world, with the safety net of keyword monitoring to protect against the risks (Impero Education, 2016).

In this process, educational discourse comes to fit closely with the evangelical vision of technology commentators such as Kevin Kelly who presented 'tracking' as one of the twelve dimensions of unavoidable technological change (Kelly, 2016).

The Digital Ocean: A New (Data-Driven) Environment for Education

The final stage of naturalisation comes when corporate educational providers imagine their interventions in and around the classroom as part of an everyday *culture* in which ubiquitous connection via mobile devices constructs 'one, seamless life' (Impero Education, 2016). In this context, corporations can imagine themselves, not as economic disruptors of the educational process, but as introducers of a 'more inclusive, *socially*-connected learning' process (Pearson, 2014a, p. 34; italics added) that is 'more interconnected, instrumented and intelligent' (IBM Education, 2014).

Pearson Education expresses this expanded vision through the environmental metaphor of 'digital ocean,' in which the use of data 'outweighs the costs' (such as privacy): '*a world in which* [because digital activities are seen as natural] *data are a side-effect, not the primary goal of interesting and motivating activity*' (Pearson, 2014b, p. 15; italics in original). The digital ocean is counter-posed to the 'digital desert' of education pre-datafication:

> We can see the digital ocean of data slowly rising from our post on the edge of a historical era we call the "digital desert." In the digital desert, data collection and storage was expensive, limited, and isolated . . . [with] no systematic large-scale way to monitor outcomes . . . The absence of mobile computing devices and information networks in the digital desert inhibited the movement and comparison of

> data across social situations and groups . . . [Conversely] the data of the digital ocean is not simply more data as we knew it in the pre-digital era . . . It is ubiquitous (coming from all manner of activity) and persistent, and it reflects social connection (Pearson, 2014b, pp. i, 1–2, 6).

Note that, in line with the previously noted alibi about data that it becomes meaningful and valuable *only through use* (excusing any issues raised by data collection), data are presented as 'only a starting point that is necessary, but not sufficient to transform education – or any activity' (Pearson, 2014b, p. i).

It is by accumulating and coordinating information from 'multiple' sources, according to Pearson, that the digital ocean becomes possible. On this view, individuals acquire an ethical responsibility to *submit* to the imperatives of data-collecting and data-processing, and the emerging connective environment of dataveillance. Any observation and inference of social processes made through datafication comes, by a strange circulate logic, to be seen as 'naturalistic':

> This emerging digital ocean, when combined with appropriate analysis and standards for use, opens the door to *new types of naturalistic observation and inference* that could help us to understand and improve ourselves. Activities and individuals in the digital ocean are intertwined in such a way that we are no longer looking at the performance of a single person in isolation in a sterile environment but, rather, at interactions between individuals, often in scenarios much closer to those found in "real life" . . . Given that learning is a social endeavour situated in particular contexts, being able to capture information that includes these interactions will allow

> us to obtain better data about learning as it occurs (Pearson, 2014b, p. 9, italics added).

The result, it is claimed, is an educational process that 'mimics' students' '"always-on, available anywhere" lifestyle' (Gibbs-Poe, 2014) and matches their 'natural expect[ation of] institutions to have already embraced this [datafied] way of working' (IBM Education, 2012b).

This voicing of what the subjects of datafication *supposedly* want is also a feature of the general discourse about Big Data and its application (WEF, 2015, p. 91). The presumption that adults can be relied on to speak on behalf of children has, however, been deconstructed in debate about the UN Convention on the Rights of the Child, since their interests do not necessarily align (Lee, 2001, p. 93). Yet, there is little evidence presented here of what sort of education young people actually want, datafied or otherwise, beyond the repetition of clichés about the much-criticised notion of 'digital natives' (Palfrey & Gasser, 2008; Tapscott, 2009).

By projecting the norms of institutions and educational professionals onto the young subjects of educational surveillance regimes, the premature conclusion is reached that student norms have *already* changed, and so students have *already* consented to – indeed now demand – the profound transformation of educational spaces by datafication:

> *Today's students expect*: Real-time connection in their classes and to each other (Blackboard; italics in original).

> It's no secret; young people are more tech-savvy than ever . . . students naturally expect access to a modern digital environment throughout their school lives (Impero Education).

This is how the transfer of educational authority to corporations can be presented as a transfer of power to the *users themselves*, even though they are students being monitored under conditions they do not control.

Digital technology has long been framed as 'disruptive' in education (Selwyn, 2015b), yet rarely has it disrupted the social relations of education itself. But beneath the rhetoric of democratising education through continuous data collection today lies a disguised threat to the *social* process of supported human development that since the early 20th century has been seen to be core to education itself (Dewey, 1971).

Conclusion: Education as Rehearsal for a Fully Datafied World

In this article, through an analysis of the discourse of eight US and UK corporate providers of educational platforms and software, we have uncovered the foundations of what Bowker and Star (1999) call a 'trajectory of naturalization' (p. 299) that helps us accept intense processes of dataveillance in educational institutions. Our argument is not based on empirical work in classrooms themselves, where important contradictions and forms of resistance may be uncovered. We have rather been concerned with unpacking an authoritative general discourse that is today a resource for framing and shaping local educational practice. This is useful, since any potential transformation of education requires a discourse to authorise and frame it.

From the simple starting-point of treating data as just naturally there, technologies of data collection and processing become, in the discourse we uncovered, treated as natural within a vision of a naturally connected environment. There follows a deeper reimagining of the educational process itself – its spaces, times and forms of agency – with new educational actors being corporations that distance themselves from an older classroom-based model of education. Educational 'visibility' is reimagined on a scale to which the teacher's eyes and ears could never have aspired. The resulting surveillant environment (Andrejevic & Selwyn, 2019), far from being a problem, is presented as the basis of a new educational value, while the agency of learners and teachers in the classroom becomes ancillary to system demands for collecting and processing data within what Pearson call the 'digital ocean.' Here, the panoptic possibilities *themselves* in classrooms are not always disguised, but the potential tensions they bring about at a deeper level of human values are given less prominence as those possibilities are treated as enabling an unquestionably better and much-needed foundation for educational objectives. In the discourse of this reimagined environment of education, older aspects of education are thus being 'forgotten' (Bowker and Star, 1999, p. 299) through the warranted, recurring use of new technologies of data collection and processing that are seamlessly incorporated into education, leaving little room for questioning the 'nature' (rather than purpose or utility) of data technologies *per se*.

There are many directions in which the implications of our argument for contemporary education could be developed further. One relates to social governance. The educational world is re-described around the agency of software developers and the marketers of digital educational platforms, who can 'nudge' (Thaler & Sunstein, 2009) higher educational performance, a role that only a decade ago would have been seen as the exclusive responsibility of teachers. Today's 'anywhere-anytime access' to

educational domain via ‘any device’ risks developing into a continuous mode of environmental influence that Karen Yeung calls the ‘hypernudge’ (Yeung, 2017), which *modulates* rather than merely monitors behaviour. Yet, as Yeung argues, the hypernudge is deeply incompatible with any notion of freedom of choice. If so, the new datafied vision of education is potentially in tension with the very model of liberal education that its purveyors supposedly espouse. The resulting growth in surveillance risks ‘chilling’ young people’s capacity for self-development, since it does not only track them, but actively interferes with their formation of will through choice (Cohen, 2012; Frischmann & Selinger, 2018; Taylor, 2017).

The developments we have discussed are not tied exclusively to schools. A broader environment is emerging today in which children, for example through their toys, ‘are socialised into surveillance culture’ (Mascheroni, 2018, p. 519), with potentially negative consequences for the fundamental human values, such as autonomy, with which surveillance remains incompatible. It has become ever more urgent to uncover the processes whereby this broader transformation is becoming naturalised in powerful discourses by educational providers and representatives of global business, as we did in this article. The next step is to challenge directly these discourses as they work to transform the world of education beyond recognition.

References

Agre, P. E. (1994). Surveillance and capture. *The Information Society*, *10*(2), 101–127.

Amoore, L. (2013). *The Politics of Possibility*. Durham, UK: Duke University Press.

Andrejevic, M., & Selwyn, N. (2019). Facial recognition technology in schools. *Learning, Media and Technology*, [online]. Retrieved from: https://doi.org/10.1080/17439884.2020.1686014

Bernstein, B. (2001). From pedagogies to knowledges. In A. Morais, I. Neves, B. Davies, & H. Daniels (Eds.), *Towards a sociology of pedagogy* (pp. 363–368). New York, NY: Peter Lang.

Blackboard. (http://www.blackboard.com/)

Blackboard. (2016). *BIG DATA HAS ARRIVED*. Retrieved from: http://www.blackboard.com/resources/pdf/2016_dataanalyticsreport_blackboard_interactive.pdf

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

Bowker, G. C. (2005). *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting Things Out*. Cambridge, MA: MIT Press.

Breiter, A., & Jarke, J. (2016, April). *Datafying education*. Bremen: Communicative Figurations. https://www.kommunikative-figurationen.de/fileadmin/user_upload/Arbeitspapiere/CoFi_EWP_No-11_Breiter_Jarke.pdf

Casella, R. (2003). The False Allure of Security Technologies. *Social Justice*, *30*(3), 82–93.

Chan, A. S. (Forthcoming). Venture Ed. In D. Ribes & J. Vertesi (Eds.), *Handbook of Digital STS*. Princeton, NJ: Princeton University Press.

Clarke, R. (1988). Information Technology and Dataveillance. *Communications of the ACM*, *31*(5), 498–512.

Cohen, J. (2012). *Configuring the Networked Self*. New Haven, CT: Yale University Press.

Couldry, N., & Yu, J. (2018). Deconstructing Datafication's Brave New World. *New Media and Society*, *20* (12), 4473–4491.

Deakin, J., Taylor, E., & Kupchik, A. (Eds.) (2018). *The Palgrave International Handbook of School Discipline, Surveillance and Social Control.* Basingstoke: Palgrave Macmillan.

Dewey, J. (1971). My Pedagogic Creed. In M. S. Dworkin (Ed.), *Dewey on Education* (pp. 19–32). New York, NY: Teachers College Press.

Dewey, J. (1977). The Relation of Theory to Practice in Education. In J. A. Boydston (Ed.), *The Middle Works of John Dewey, vol.3: 1903-1906* (pp. 249–272). Carbondale, IL: Southern Illinois University Press.

Eynon, R. (2013). The Rise of Big Data: what does it mean for education, technology and media research? *Learning, Media and Technology*, *38*(3), 237–240.

Frase, K. (2014). Big Data: Connecting Education and Economic Development. Retrieved from: https://www.ibm.com/blogs/insights-on-business/education/big-data-connecting-education-and-economic-development/

Frischmann, B., & Selinger, E. (2018). *Re-engineering Humanity*. New York, NY: Cambridge University Press.

Gallagher, M. (2010). Are Schools Panoptic? *Surveillance & Society*, *7*(3/4), 262–272.

Gibbs-Poe, H. (2014). Delivering Value in K-12 with Mobile, Social, and Analytics. Retrieved from: https://www.ibm.com/blogs/insights-on-business/education/delivering-value-in-k-12-with-mobile-social-and-analytics/

Gillespie, T. (2014). The Relevance of Algorithms. In T. Gillespie, P. J. Boczkowski, & K. A. Foot (Eds.), *Media Technologies* (pp. 167–194). Cambridge, MA: MIT Press.

Gilliom, J., & Monahan, T. (2013). *SuperVision*. Chicago, IL: University of Chicago Press.

Gioga, M. (2016). Social Media Insights: Classroom of the Future. Retrieved from: https://www.ibm.com/blogs/insights-on-business/education/social-media-insights-education/

Giroux, H. (2015). *Education and the Crisis of Public Values*. New York, NY: Peter Lang.

Gitelman, L. (Ed.) (2013). *"Raw Data" is an Oxymoron*. Cambridge, MA: MIT Press.

Hill, A. (2014, September 15). A day in the Life of a Data mined Kid. *Marketplace*. Retrieved from: https://www.marketplace.org/2014/09/15/education/learning-curve/day-life-data-mined-kid

IBM. (2009). *Building a Smarter Classroom*. Retrieved from: ftp://ftp.software.ibm.com/software/data/sw-library/cognos/pdfs/solutionbriefs/sb_building_a_smarter_classroom_solution_brief.pdf

IBM. (2016). First IBM Watson Education App for iPad Delivers Personalized Learning for K-12 Teachers and Students. Retrieved from: http://www-03.ibm.com/press/us/en/pressrelease/50815.wss

IBM Education. (2012a). *IBM Classroom in the Cloud.* Retrieved from: https://www-935.ibm.com/services/multimedia/IBM_Classroom_in_the_Cloud.pdf

IBM Education. (2012b). *IBM Exceptional Student Experience*. Retrieved from: https://www-935.ibm.com/services/multimedia/Achieving_an_Exceptional_Student_Experience.pdf

IBM Education. (2014). *The future of learning*. Retrieved from: https://afd34ee8b0806295b5a7-9fbee7de8d51db511b5de86d75069107.ssl.cf1.rackcdn.com/EDW03005USEN.PDF

IBM Education. (2016). *Personalised education*. Retrieved from:

https://www.ibm.com/thought-leadership/technology-market-research/personalised-education-quiz/dist/files/ibm-white-paper.pdf

IBM Watson Education. (https://www.ibm.com/watson/education/)

Impero Education. (https://www.imperosoftware.com/)

Impero Education. (2016). *White paper*. Glastonbury, CT: Digital Citizenship Institute. Retrieved from:

http://www.learningnetwork.ac.nz/shared/professionalReading/DIGCIT.pdf

Jenks, C. (Ed.) (2005). *Childhood*. New York, NY: Routledge.

Kelly, K. (2016). *The Inevitable.* New York, NY: Penguin.

Kitchin, R. (2014). Big Data: new epistemologies and paradigm shifts. *Big Data and Society, 1*(1), 1–12.

Knewton. (https://www.knewton.com)

Knewton. (2014). *Heavy Duty Infrastructure for the Adaptive World*. Retrieved from:

https://www.knewton.com/assets-v2/downloads/knewton-intro-2014.pdf

[currently unavailable]

Lewis, S., & Holloway, J. (2019). Datafying the teaching 'profession'. *Cambridge Journal of Education*, *49*(1), 35–51.

Lewis, T. (2003). The Surveillance Economy of post-Columbine school. *Review of Education Pedagogy and Cultural Studies*, *25*(4), 335–356.

Lindh, M., & Nolin, J. (2016). Information We Collect. *European Educational Research Journal*, *15*(6), 644–663.

Lindh, M., & Nolin, J. (2017). GAFA speaks. *Journal of Documentation*, *73*(1), 160–180.

Nolin, J. (2019). Data as oil, infrastructure or asset? Three metaphors of data as economic value. *Journal of Information, Communication and Ethics in Society*, [online]. Retrieved from: https://doi.org/10.1108/JICES-04-2019-0044

Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. (2016). *Intelligence Unleashed.* London: Pearson/UCL Knowledge Lab.

Lupton, D., & Williamson, B. (2017). The datafied child. *New Media and Society*, *19* (5), 780–794.

Mascheroni, G. (2018). Researching Datafied Children as Datafied Citizens. *Journal of Children and Media, 12*(4), 517–523.

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data*. London: John Murray.

Mayer-Schönberger, V., & Cukier, K. (2014). *Learning with Big Data*. New York, NY: Houghton Mifflin Harcourt.

OECD. (2015). *Data-Driven Innovation*. Paris: OECD Publishing.

O'Neil, C. (2016). *Weapons of Math Destruction*. London: Allen Lane.

Palfrey, J., & Gasser, U. (2008). *Born Digital*. New York, NY: Basic Books.

Pearson. (https://www.pearson.com/)

Pearson. (2014a). *A Rich Seam: How New Pedagogies Find Deep Learning*. Retrieved from: https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/ARichSeamEnglish.pdf

Pearson. (2014b). *Impacts of the Digital Ocean on Education*. Retrieved from: https://www.pearson.com/content/dam/one-dot-com/one-dot-com/global/Files/about-pearson/innovation/open-ideas/DigitalOcean.pdf

Pearson. (2014c). *Preparing for a Renaissance in Assessment*. Retrieved from:

https://www.pearson.com/content/dam/one-dot-com/one-dot-com/uk/documents/educator/primary/preparing_for_a_renaissance_in_assessment_and_summary_text_december_2014.pdf

Potter, J. (1990). *Representing Reality.* Thousand Oaks, CA: Sage.

Saltman, K. (2014). Neoliberalism and Corporate School Reform. *Review of Education, Pedagogy and Cultural Studies, 36*(4), 249–259.

Sandvig, C., Hamilton, K., Karahalios, K., & Langbort, C. (2016). When the Algorithm Itself Is a Racist. *International Journal of Communication, 10*, 4972–4990.

Selwyn, N. (2014). *Digital Technology and the Contemporary University*. Oxon: Routledge.

Selwyn, N. (2015a). Data Entry: Towards the Critical Study of Digital Data and Education. *Learning, Media and Technology*, *40*(1), 64–82.

Selwyn, N. (2015b). The discursive construction of education in the digital age. In R. H. Jones, A. Chik, & C. A. Hafner (Eds.), *Discourse and Digital Practices* (pp. 226–240). London: Routledge.

Selwyn, N. (2016a). *Education and Technology* 2nd ed. London: Bloomsbury.

Selwyn, N. (2016b). *Is Technology Good for Education?* Cambridge, UK: Polity.

Siemens, G., & Gasevic, D. (2012). Guest Editorial – Learning and Knowledge Analytics. *Educational Technology and Society*, *15*(3), 1–2.

Tapscott, D. (2009). *Grown Up Digital*. New York, NY: McGraw-Hill.

Taylor, E. (2013). *Surveillance Schools*. London: Palgrave Macmillan.

Taylor, E. (2017). 'If I wanted to be on Big Brother, I would've auditioned for it'. In E. Taylor & T. Rooney (Eds.), *Surveillance Futures* (pp. 19–35). London: Routledge.

Taylor, E. (2018). Student drug testing and the surveillance school economy. *Journal of Education Policy*, *33*(3), 383–397.

Taylor, E., & Rooney, T. (2017). Digital playgrounds. In E. Taylor & T. Rooney (Eds.), *Surveillance Futures* (pp. 1–16). London: Routledge.

Thaler, R. H., & Sunstein, C. R. (2009). *Nudge*. New Haven, CT: Yale University Press.

United Nations. (2012). *Big Data for Development*. Retrieved from: http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseMay2012.pdf

Van Brakel, R. (2016). Pre-emptive Big Data Surveillance and its (Dis)empowering Consequences. In B. v. d. Sloot, D. Broeders, & E. Schrijvers (Eds.), *Exploring the Boundaries of Big Data* (pp. 117–141). Amsterdam: Amsterdam University Press.

Van Dijck, J. (2014). Datafication, dataism and dataveillance. *Surveillance and Society*, *12*(2), 197–208.

Van Dijck, J., Poell, T., & De Waal, M. (2018). *The Platform Society*. Oxford: Oxford University Press.

White House. (2014). *Big Data and Privacy*. Retrieved from: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/PCAST/pcast_big_data_and_privacy_-_may_2014.pdf

Williamson, B. (2015). Governing software. *Learning, Media and Technology*, *40*(1), 83–105.

Williamson, B. (2017). Calculating Children in the Dataveillance School. In E. Taylor & T. Rooney (Eds.), *Surveillance Futures* (pp. 50–66). London: Routledge.

World Economic Forum. (2011). *Personal Data: The Emergence of a New Asset Class*. Retrieved from: http://www3.weforum.org/docs/WEF_ITTC_PersonalDataNewAsset_Report_2011.pdf

World Economic Forum. (2012a). *Rethinking Personal Data: Strengthening Trust*. Retrieved from: http://www3.weforum.org/docs/WEF_IT_RethinkingPersonalData_Report_2012.pdf

World Economic Forum. 2012b. *The Global Information Technology Report 2012*. Retrieved from: http://www3.weforum.org/docs/Global_IT_Report_2012.pdf

World Economic Forum. 2013. *Unlocking the Value of Personal Data*. Retrieved from: http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf

World Economic Forum. 2015. *The Global Information Technology Report 2015*. Retrieved from: http://www3.weforum.org/docs/WEF_Global_IT_Report_2015.pdf

Yeung, K. (2017). 'Hypernudge': Big Data as a mode of regulation by design. *Information, Communication and Society*, *20*(1), 118–136.

[1] Our approach will focus primarily on the 'discourse' of educational platform providers. For a parallel critical approach that focusses on broadly organisational convergences in education from which platforms benefit, see Van Dijck, Poell, & De Waal (2018).

[2] There are potential parallels here with developments in areas such as policing (van Brakel, 2016) and territorial control (Amoore, 2013).

[3] Such sources with regularly changing websites and quotations are cited without date, unless referring to a distinct document.

[4] In this sense, we were inspired by some versions of Discourse Analysis (Potter, 1990), though our study is not formally a discourse analysis. Compare for a parallel strategy looking for absences as well as presences in educational discourse, Lindh and Nolin (2016) on Google Apps for Education.

[5] For more detail, see Couldry and Yu (2018).

[6] For further discussion on metaphors of data use, see Lindh and Nolin (2017); Nolin (2019).