1

INTRODUCTION

Learning machines, digital data and the future of education

In October 2015, over 1,000 young software developers and hackers attended HackingEDU, a three-day educational hackathon held at the San Mateo Event Center in San Francisco. Originally launched at the 2014 Google Summit, the annual HackingEDU event – the 'world's largest educational hackathon' – is intended to help software developers and programmers, most of them college students, 'revolutionize the education industry' while competing for over US$100,000 in prizes (Hunckler, 2015). Featuring expert workshops, panel discussions and guest speakers, HackingEDU 2015 was supported by major technology companies including IBM, Google, Uber, PayPal and Automattic, as well as by successful educational technology businesses such as Chegg and EdModo. It emphasized the ways in which technologies might be used to 'disrupt' and 'revolutionize' education, much as 'Uber revolutionized the transportation industry based on a simple concept: press a button, get a ride', as the event's partnership director phrased it (Uber 2015). The technology projects produced during HackingEDU 2015 included titles such as Learnization, CereBro, PocketHelp, QuizPrep, BrainWars and StudyTracker, almost all of them relying on a combination of digital data and database technologies and constructed by their young designers using a variety of programming languages, software programs and hardware devices.

Elsewhere in San Francisco, many other fledgling edtech projects are annually developed through the support of edtech 'incubator' or 'accelerator' programs. Incubators typically help entrepreneurs and new startups to test and validate ideas, while accelerators turn products into scalable businesses, often through direct equity investment, and help provide entrepreneurs with legal, IT and financial services along with mentorship, working space and access to educators, entrepreneurs, business partners and potential investors (Gomes 2015). For example, Imagine K12 is 'a startup accelerator focused on education technology':

> Our goal is to improve your company's chances of success. We do this through a combination of strategic advice and mentorship, a series of speakers and seminars designed to help founders make better decisions, value-added networks of entrepreneurs and educators, and $100,000 of initial funding. … Companies begin receiving support from Imagine K12 immediately upon their acceptance, including $20k of funding. … [A]ll accepted startups are required to move to Silicon Valley for an intensive four-month program. (Imagine K12 2015)

Edtech incubator and accelerator programs like Imagine K12 provide the space, support and investment required for programmers to write educational technologies, and ultimately act as mechanisms that might realize the 'revolutionary' ambitions of entrants to competitions like HackingEDU. Notably, Imagine K12 has since merged with another accelerator program, Y Combinator, an organization established by billionaire PayPal founder Peter Thiel, a major donor and spokesperson during Donald Trump's US presidential campaign in 2016.

A key educational technology advocate, Thiel has supported and funded many companies and startups that focus on 'revolutionizing' education through data-driven software applications (Levy 2016). For new startups that successfully graduate from the incubation and acceleration stage, entrepreneurial investors from Silicon Valley have been funding educational technology projects with unprecedented financial enthusiasm since about 2010 (EdSurge 2016). With webs of political support and entrepreneurial investment for educational technology growing, a new digital future for education is being imagined and pursued in governmental and private sector settings alike, with significant consequences for learning, policy and practice.

HackingEDU is an important event with which to start this book for a number of reasons. It locates education as it currently exists as a problematically broken system which is in need of revolutionizing. It proposes that the solution is in the hands of software developers and hackers who can write code. It suggests that the availability of masses of educational data can be used to gain insights into the problems of education, and to find solutions at the same time. And it also demonstrates how private sector technology companies have begun to fixate on education and their own role in fixing it. Incubators and accelerators such as Imagine K12 and Y Combinator can then step in with entrepreneurial experience to grow new products into successful startup businesses, to enable programmers to fine-tune the code and algorithms required to make their product run, and to gain financial investment required to push it out into practice. The promise appears simple. Take a model like Uber, the mobile app that has transformed taxi services by harvesting locational data from its millions of users, and then translate that model into a template for educational reform. Fund, incubate and accelerate it until it performs optimally. All it takes to revolutionize education for the future is a few million lines of software code and big piles of digital data.

Digitizing and Datafying Education

The goal of this book is to understand and detail how digital data and the code and algorithms that constitute software are mixing with particular political agendas, commercial interests, entrepreneurial ambitions, philanthropic goals, forms of scientific expertise, and professional knowledge to create new ways of understanding, imagining and intervening in education. Education is now a key site in which big data and algorithmic techniques of data mining and analysis performed with software are proliferating and gaining credibility.

Yet the quantitative increase in data brought about by recent developments and the qualitative effects they are beginning to exert in education have gone largely unnoticed amid much more high-profile concerns about the data mining

conducted by social media companies on their users, targeted online advertising that is driven by consumer data, or the data-based forms of surveillance being practised by governments (van Dijck 2013). A 'new apparatus of measurement has drastically expanded' with the availability of digital data in diverse areas of public and private life, 'allied with a set of cultural changes in which the pursuit of measurement is seen to be highly desirable' (Beer 2016a: 3). Education, by contrast, appears more 'ordinary':

> Given that so much attention has already been paid to social media corporations and governmental and security agencies, what we now need to attend to is other, more ordinary actors, as social media data mining becomes ordinary. (Kennedy 2016: 7)

This book takes up the challenge of investigating the digital data technologies, organizations and practices that are increasingly becoming integrated into many aspects of education. A vast apparatus of measurement is being developed to underpin national educational systems, institutions and the actions of the individuals who occupy them.

While the pursuit of educational measurement has a long history stretching back to the nineteenth century (Lawn 2013), it is being extended in scope, enhanced in its fidelity, and accelerated in pace at the present time as new technologies of big data collection, analysis and feedback are developed and diffused throughout the system (Beneito-Montagut 2017; Selwyn 2015). Similarly, schools, colleges and universities have employed e-learning programs for many years in their pedagogic and instructional processes (Selwyn 2011), but with big data and analytics processes now increasingly augmenting them, these resources can now adapt to their users and 'talk back' to educators (Mayer-Schönberger and Cukier 2014). Software and digital data are becoming integral to the ways in which educational institutions are managed, how educators' practices are performed, how educational policies are made, how teaching and learning are experienced, and how educational research is conducted.

The presence of digital data and software in education is being amplified through massive financial and political investment in educational technologies, as well as huge growth in data collection and analysis in policymaking practices, extension of performance measurement technologies in the management of educational institutions, and rapid expansion of digital methodologies in educational research. To a significant extent, many of the ways in which classrooms function, educational policy departments and leaders make decisions, and researchers make sense of data, simply would not happen as currently intended without the presence of software code and the digital data processing programs it enacts.

To fully appreciate how digital data are being generated and exerting material effects in education, then, it is essential to view data and the software code and algorithms that process it in relation to a range of other factors that frame their use.

Political agendas relating to education policy and governance, commercial interests in the educational technology market, philanthropic and charitable goals around supporting alternative pedagogic approaches, emerging forms of scientific expertise such as that of psychology, biology and neuroscience, as well as the practical knowledge of educator professionals, all combine with new kinds of data practices and digital technologies. That is, the mobilization of digital data in education happens in relation to diverse practices, ways of thinking, ambitions, objectives and aspirations that all shape how data is put to use, define the tasks and projects through which data is deployed, and co-determine the results of any form of educational data analysis. The role and consequences of digital data in education cannot be understood without appreciating their relations with the other ordinary features of education – policies, accountability mechanisms, commercial imperatives, charitable intentions, scientific knowledge and professional practice.

In this sense, the subject of this book is the combined process of 'datafying' and 'digitizing' education. Putting it simply, 'datafication' refers to the transformation of different aspects of education (such as test scores, school inspection reports, or clickstream data from an online course) into digital data. Making information about education into digital data allows it to be inserted into databases, where it can be measured, calculations can be performed on it, and through which it can be turned into charts, tables and other forms of graphical presentation. 'Digitization' refers to the translation of diverse educational practices into software code, and is most obvious in the ways that aspects of teaching and learning are digitized as e-learning software products. If you want to build some digital e-learning software, you have to figure out how to do that in lines of code: to encode educational processes into software products. Diverse aspects of education from policy, leadership, management and administration to classroom practice, pedagogy and assessment are now increasingly subjected to processes of digitization, as software is coded and algorithms are designed to augment and rework everyday tasks and processes across the education sector.

Datafication and digitization support and complement one another in myriad ways. For example, when a piece of e-learning software is coded in digital form, it is often designed in such a way that it can generate information about the ways that it is used (visible in, for example, the log files that demonstrate how a user has interacted with the software). That information can then be used, as analysable digital data, to help the producers of the software learn more about the use of their product, data which can then be used to help inform the writing of better code (a software patch, upgrade or update) or the programming of new software products altogether. To take another example: when millions of learners around the world all take a standard global test, the activities they undertake ultimately contribute to the production of a massive database of test results. Making sense of the vast reserves of data in such a

database can only be accomplished using software that has been coded to enable particular kinds of analyses and interpretations. The software does not have to be especially appealing – the datafication of education depends to a significant degree on the digital coding undertaken to produce very mundane software products like spreadsheets and statistical analysis packages – but it is certainly becoming more seductive with the ready availability of highly graphical forms of data visualization software, as well as more accessible and easier to use. With both educational technologies and educational data, processes of digitization and datafication support and reinforce each other.

In short, much of education today is being influenced and shaped by the production of lines of code that make digital software function, and by the generation of digital data that allows information about education to be collected, calculated and communicated with software products. Does this matter? Yes, it matters urgently, because the coding of software products for use in education, or the application of coded devices that can process educational digital data, are beginning to transform educational policies, pedagogies and other practices in ways which have so far been the subject of very little critical attention.

As new kinds of software are developed for use in educational contexts that rely on both software code and digital data, we are beginning to see new ways in which schools, universities, educational leaders, teachers, students, policymakers and parents are influenced. Schools are being turned into data-production centres, responsible for constantly recording and auditing every aspect of their performance (Finn 2016). Leaders are being called on to act on their data to improve the institutions they manage (Lewis and Hardy 2016), often using 'learning management systems' to assist in administrative tasks (Selwyn et al. 2017). Students are becoming the subjects of increasingly pervasive data mining and data analytics packages that, embedded in educational technologies and e-learning software, can trace their every digital move, calculate their educational progress and even predict their probable outcomes (Suoto-Otero and Beneito-Montagut 2016). Students in universities are experiencing ever-greater use of online tools to measure their progress (Losh 2014), with their assignments being entered into massive global plagiarism detection databases (Introna 2016). At the same time, university managers are required to make use of complex performance indicator metrics and institutional data dashboards to facilitate decision-making and planning (Wolf et al. 2016). Even early years settings such as nurseries are increasingly required to collect data on young children's development so that it can be tracked against national and international benchmarks (Roberts-Holmes 2015; Moss et al. 2016), which is mirrored by the growing use of analytics technologies in adult education and professional learning (Fenwick and Edwards 2016).

Beyond the spaces of learning, policymakers are increasingly exhorted to develop data-driven or 'evidence-based' policies that are crafted in response to insights derived from digital data (Sellar 2015a), including school inspection

data presented on institutions' 'data dashboards' (Ozga 2016). Parents, too, are encouraged to become educational data analysts who use digital 'school comparison' websites to inform their choices about which schools to enrol their children in (Piattoeva 2015). For teachers, a new industry in educational 'talent analytics', or 'labour market analytics', has even appeared (Beneito-Montagut 2017), with fully-automated software products like TeacherMatch acting as 'advanced education talent management' platforms for the recruitment, assessment, professional development and 'talent investment' of teachers, using matching algorithms to match schools with staff just like a social media dating service (TeacherMatch 2015).

Many commercial organizations are changing their business models and practices to engage in education, such as Google with its Google Apps for Education suite of free-to-use cloud services for schools (Lindh and Nolin 2016). Meanwhile, existing commercial 'edu-businesses' such as Pearson – a global education textbook publisher – have moved to become prominent educational software providers and key collectors of educational data (Hogan et al. 2015). Commercial tools for data collection, processing and analysis are finding their way into the discipline of educational research, knowledge production and theory generation too, in ways that are reshaping how education is known and understood (Cope and Kalantzis 2016). And finally, an increasing number of private sector 'data brokers' are starting to collect education-related data, curate and aggregate it using analytics tools, and sell it back to education stakeholders (Beneito-Montagut 2017).

It's not just the people and organizations of education that are affected by the recent acceleration of data-processing software, but curriculum, pedagogy and assessment too. The notion of a curriculum containing the content-knowledge to be taught in schools is itself being challenged, as new kinds of 'adaptive' learning software are developed that can semi-automate the allocation and 'personalization' of content according to each learners' individual data profile (Bulger 2016). Pedagogy is being distributed to automated machines such as 'teacher bots' and 'cognitive tutors': computerized software agents designed to interact with learners, conduct constant real-time analysis of their learning, and adapt with them (Bayne 2015). And the notion of assessment as a fixed event is being supplanted by real-time assessment analytics and computer-adaptive testing, which automatically assess each learner on-the-go and adapt to their responses in real-time (Thompson 2016). What is even meant by 'learning' is being questioned with the collection of datasets so large that enthusiasts believe they can reveal new truths about learning processes that educational researchers working within disciplinary frameworks such as psychology, sociology and philosophy have been unable to detect before (Behrens 2013).

Many of these developments and innovations with digital software and data in education exist technically, but they are also the product of extensive claims, promotional activity and imaginative marketing which centres on the idea that

technical solutions have the capacity to transform education for the future. Businesses with products to sell, venture capital firms with return on investment to secure, think tanks with new ideas to promote, and policymakers with problems to solve and politicians with agendas to set have all become key advocates for data-driven education. Of course, we need to be at the very least cautious about many of the claims made about the transformative and revolutionary potential of many new developments, if not downright sceptical – and, indeed, a little resistant.

But the point I pursue throughout is that what we are currently witnessing are signs of a new way of thinking about education as a datafied and digitized social institution. Seriously powerful organizations are at work in this space, organizations with a forceful and influential shared imagination concerning the future of education. It is easy to be dismissive of the claims-making, hype and hubris that surround emerging developments like learning analytics and computer-based cognitive tutors. But it's less easy to dismiss these developments and the claims that support them when you can see that some of the world's richest and most powerful companies are dedicating extraordinary research and development resources to them; when you can read reports advocating and sponsoring them by influential think tanks; when you hear that politicians are backing them; when you discover that enormous sums of venture capital and philanthropic funding are being invested to make them a reality.

A shared vision of the digitization and datafication of education is emerging. Diverse ideas and actors have combined to produce collective imaginative resources that can be used to animate research and development (R&D) practices, to persuade politicians, to generate investment, and to galvanize new practices (Jasanoff 2015). Of course, education has long been a site of future imagination. A 'dominant myth of the future of education' in recent years has been one that 'emerges out of an instrumental conception of education as primarily concerned with serving the formal economy' (Facer 2011: 8). Visions of data-driven education complicate this dominant myth of the future. While economic fantasies of human capital development persist, they are being supplemented and extended by dreams of new forms of governance and citizenship, new scientific aspirations of psychological optimization and cognitive enhancement, and new commercial objectives to insert private sector technologies and practices into public education.

Myths and imaginative visions, moreover, can become material realities when given technical form and inserted into social contexts. The developments traced out in the following chapters are all parts of a new emerging imaginary of the digital future of data-driven education that appears to be considered desirable, and that many organizations and individuals seem to agree could and should be attained through putting new technical developments into practice in the present. The twin processes of digitization and datafication form the basis for the book, but the practices of coding educational technologies of various kinds and of

datafying education through diverse techniques are all also situated contextually and are animated by a particularly powerful imaginative resource which envisions education as a massively data-driven and software-supported social institution. The difference that digital data make in education is the result of the highly diverse efforts of programmers, project managers, businesses, startup accelerator programs, policymakers and politicians, think tanks and innovation labs, school managers, leaders, and educators themselves – the material practices of all of them shaped by an imagined vision of a digitized and datafied future which has become increasingly pervasive, persuasive and seemingly desirable.

Datafying Education

'Datafication' refers to the transformation of many aspects of education into quantifiable information that can be inserted into databases for the purposes of enacting different techniques of measurement and calculation. Datafication itself has a long history, detailed more fully in Chapter 2. Recent developments such as the establishment of data labs and data centres for educational data mining and analysis, and the proliferation of specific products such as learning analytics, adaptive learning software and computerized tutors, all rely on the constant collection of masses of digital data. Large-scale educational data has been available from the aggregation of test results or school census information for decades. The key shift with big data is that it is now collected in or near real-time directly as learners interact with software systems. That is to say, large-scale datasets have been historically gathered primarily through assessments and data collection events that have to be separated off from the normal rhythms of the classroom; big data are captured from within the pedagogic machinery of the teaching and learning process itself by being pieced together from the millions of data points that are generated as learners click on content and links, engage with digital educational materials, interact with others online, and post responses to challenges. Digital course content, online courses, e-textbooks, digital simulations, and more, provide the front-end interface for the production of educational big data, behind which lies a sophisticated back-end infrastructure of data collection, information storage, algorithmic processing, and analytics and data visualization capacities.

Underlying these developments is a set of powerful animating visions or imaginaries of datafication. The authors of *Learning with Big Data: The Future of Education* (Mayer-Schönberger and Cukier 2014) imagine that big data will 'reshape learning' through 'datafying the learning process' in three significant ways: (1) through real-time feedback on online courses and e-textbooks that can 'learn' from how they are used and 'talk back' to the teacher; (2) individualization and personalization of the educational experience through adaptive

learning systems that enable materials to be tailored to each student's individual needs through automated real-time analysis; and (3) probabilistic predictions generated through data analytics that are able to harvest data from students' actions, learn from them, and generate predictions of individual students' probable future performances. The authors imagine school as a 'data platform' where the real-time datafication of the individual is becoming the 'cornerstone of a big-data ecosystem', and in which 'educational materials will be algorithmically customized' and 'constantly improved' (Mayer-Schönberger and Cukier 2014).

A significant amount of data-driven activity has been undertaken in the higher education sector, through widespread use of learning management systems and online programs such as MOOCs (Massive Open Online Courses) (Knox 2016). But schools are also being targeted for datafication. The US think tank the Center for Data Innovation has produced a report advocating a vision of a 'data-driven education system' for schooling. 'U.S. schools are largely failing to use data to transform and improve education, even though better use of data has the potential to significantly improve how educators teach children and how administrators manage schools', its author claims (New 2016: 1). Instead, the think tank argues that a data-driven education system should achieve four main goals:

> *Personalization*: Educators dynamically adjust instruction to accommodate students' individual strengths and weaknesses rather than continue to utilize a mass production-style approach.
>
> *Evidence-Based Learning*: Teachers and administrators make decisions about how to operate classrooms and schools informed by a wealth of data about individual and aggregate student needs, from both their own students as well as those in comparable schools across the nation … rather than by intuition, tradition, and bias.
>
> *School Efficiency*: Educators and administrators use rich insight from data to explore the relationships between student achievement, teacher performance, and administrative decisions to more effectively allocate resources.
>
> *Continuous Innovation*: Researchers, educators, parents, policymakers, tech developers, and others can build valuable and widely available new education products and services to uncover new insights, make more informed decisions, and continuously improve the education system. (New 2016: 2)

These goals for data-driven education systems accurately capture the dominant imaginary related to the collection and use of data in schools. 'Personalization' has become perhaps the main keyword of data-driven education, emphasizing systems and processes that can be intelligently tailored to the individual students. The use of evidence to perform comparisons across institutions and systems has a long lineage in education policy, but with digitization is becoming much easier and quicker to conduct. Achieving efficiency is paramount for schools, with performance management tools now available to ensure that students, teachers and

administrators are all producing measurable outputs. And as larger and larger quantities of data become available – as masses of educational big data – new patterns and insights are being sought to address the goals of various stakeholders, such as the improvement agendas of policymakers and the new product development plans of businesses. The imagined datafication of schools is to be attained through pursing these goals of personalization, evidence-based learning, efficiency and continuous innovation.

How do such goals and imaginative visions look in practice? Compelling examples of how the datafication of schools might look in the imagined near future of education are provided by Silicon Valley 'startup schools'. Startup schools are new educational institutions designed as alternatives to the mainstream state schooling model, and they originate in the technology entrepreneurship culture of Silicon Valley, the technofinancial heart of the global tech industry. A prominent example is AltSchool, set up in 2013 by Max Ventilla, a technology entrepreneur and former Google executive. It 'prepares students for the future through personalized learning experiences within micro-school communities', and its stated aim is to 'help reinvent education from the ground up' (AltSchool 2015a). A recent profile of its founder claimed that 'when Ventilla quit Google to start AltSchool, in the spring of 2013, he had no experience as a teacher or an educational administrator. But he did have extensive knowledge of networks, and he understood the kinds of insights that can be gleaned from big data' (Mead 2016). After establishing in four sites in San Francisco as a 'collaborative community of micro-schools', AltSchool later expanded to Brooklyn and Palo Alto, with further long-term plans for new schools and partnerships across the US. It has since hired executives from Google, Uber and other successful Silicon Valley startups, many with experience of big data projects. The AltSchool chief technology officer, formerly the engineer in charge of the Google.com homepage and search results experience, has stated that 'I am highly motivated to use my decade of Google experience to enable the AltSchool platform to grow and scale' (AltSchool 2015a). The AltSchool 'platform' is described as a new 'central operating system for education', one designed according to 'technology-enabled models' that are transforming other industries and institutions, such as Uber and Airbnb (AltSchool 2015b).

The models it refers to are those of the datafication of other sectors. Airbnb represents the datafication of accommodation letting. Uber has thoroughly datafied taxi services. AltSchool has been programmed to run on the same basic model, or operating system, as these datafied sectors. Thus, it depends on a sophisticated data analytics platform. Its suite of digital tools is intended to 'make personalized education a reality', which it seeks to accomplish by supporting teachers to 'develop Personalized Learning Plans and to capture student progress toward them':

> We also create platforms for efficient classroom administration so teachers have more quality face-to-face time with their students. … To ensure we are always learning from what happens outside the classroom, we build digital tools to support collaboration between teachers, parents and students. … Our project-based education approach truly comes alive when supported by carefully curated learning tools. We mentor each student in the use of technology for learning and help them skilfully navigate today's information terrain. (AltSchool 2016)

The data platform driving AltSchool is not just a technical system: it has been constructed to support a particular cultural vision of education as being 'personalized' around each individual. Personalization is its dominant ideal, and it is personalization that has been achieved successfully within the commercial social media activities of many Silicon Valley companies. For instance, Google search results are automatically personalized to each user based on their web search history. The Facebook timeline is personalized around the friends graph it constructs about each user's social network connections. The logic of personalization drives the ways in which social media platforms make recommendations for people to follow, consumer goods to buy, memes to share and so on. The culture and techniques of personalization from the commercial social media sphere are inserted into schooling through spaces such as AltSchool, and built in to its data platforms as a technical back-end complement to the front-facing cultural vision of education it projects. AltSchool ultimately balances and assembles a range of resources that appear unproblematically to crisscross the traverse between technological ideals and educational concepts.

Beyond technical and cultural similarities with the datafying priorities of the tech industry, the startup school also enjoys the financial benefits of Silicon Valley startup culture. On its establishment, AltSchool originally raised US$33 million in venture capital funding, with another US$100 million investment in 2015, including donations from Facebook's Mark Zuckerberg and the venture capital firm Andreeson Horowitz (AltSchool 2015b). AltSchool is, then, thoroughly governed, managed and financed through the discourses and material practices of Silicon Valley startup culture. Its operating system is modelled on social media data analytics. Its funding is almost exclusively generated through venture capital and tech philanthropy. Its engineering and design team are applying their social media expertise in data dashboards, algorithmic playlisting, adaptive recommender systems and app development to the development of new personalized edtech devices and platforms. The datafication of education prototyped by AltSchool, and other startup school models, is not just a technical accomplishment but the product of a financial investment model for Silicon Valley startups that has been trialled in other sectors, transplanted into education, and appears to be on the cusp of being scaled-up as a competitive market solution to the problem of mainstream schooling. Technology visionaries and imaginative entrepreneurs like Max Ventilla are becoming high-status

education reformers, using their technical expertise in software development and data analytics, combined with the entrepreneurial business expertise required to generate investment, as powerful resources to attract others to their educational visions.

As the AltSchool example demonstrates, big data is not just technical. It is, rather, the 'manifestation of a complex sociotechnical phenomenon that rests on an interplay of technological, scientific, and cultural factors':

> While the *technological dimension* alludes to advances not only in hardware, software, but also infrastructure and the *scientific dimension* comprises both mining techniques and analytical skills, the *cultural dimension* refers to (a) the pervasive use of ICTs in contemporary society and (b) the growing significance and authority of quantified information in many areas of everyday life. (Rieder and Simon 2016: 2, italics in original)

Throughout the chapters that follow, the datafication of education is treated as the contingent materialization of future visions, technologies and skilled scientific techniques, as well as of political, commercial and philanthropic ambitions, all of which are combining into hybrid sociotechnical systems for data-driven measurement and management.

Digitizing Education

Datafication of education requires learning environments to be highly instrumented to collect information (Cope and Kalantzis 2015). This means the learning environment needs to be increasingly digitally-mediated, or digitized, as AltSchool's technical 'operating system' demonstrates. The use of the term 'digitization' refers to 'the process of converting information from analog into discrete units of data that can be more easily moved around, grouped together, and analysed' (Gregory et al. 2017: xviii) using computer technologies. With the digitization of education into information that can be processed by a computer, software and the code that enacts it becomes a significant influence in how education is organized. Software code has become a system for regulating many of the practices and processes of education, teaching and learning.

Described in more detail in Chapter 3, it is important from the outset to acknowledge that code is both a product – the end-result of the work of programmers, working in real material conditions, with their own professional cultures and values, and whose coding practices are shaped by business plans and objectives – and as a productive force in the world (Kitchin and Dodge 2011). By describing code as 'productive' registers the ways in which code is programmed to perform tasks that it then enacts (or, to use the specific computational term, 'executes'). Code instructs a software program to 'do something' on a computer,

and in that basic sense it can be seen as productive. But it is also productive because writing code to execute a particular kind of task also fundamentally alters the nature of the task it is being instructed to perform (Mackenzie 2006).

In this book I focus on the ways that turning educational things into code then loops back to change education. However, to think of code just in technical terms, as a script for instructing software written in specific programming languages, would be misleading. It is certainly the case that e-learning software, policy databases and school management programs depend on lines of code for their functioning. But that code has itself to be written, or produced, as noted earlier. Programmers have to craft it, using specific kinds of programming languages and code repositories. Those programmers work according to the business plans, project management schedules and objectives of their employers. Those business plans are the operational manifestation of powerful future visions. The code produced to make software programs function is also dependent on financial investment, funding programmes and economic priorities. This goes beyond the straightforward allocation of programmers' salaries and includes the work of entrepreneurs in securing venture capital for software startups, of politicians providing tax incentives for technology companies, and of philanthropists making donations to finance new technical innovations. The software programs that enact much of education today, in other words, are also the product of imaginative business and political programmes.

An illustrative example of how digital imaginaries, software, finance and politics are interwoven in the contemporary transformation of education is provided by Edtech UK. This organization is 'a new strategic body set up to help accelerate the growth of the UK's education technology sector in Britain and globally':

> the new body is a 'front door' for industry, investment and government and a convening voice for all of the education and learning technology sector including educators, startups, scale up and high growth companies, large corporations, investors, regulators and policy makers. The focus of Edtech UK is to help support, showcase and develop the sector, with a focus on creating more jobs, developing new skills, understanding what works and driving economic growth. Its focus will be global from the outset with an ambitious programme of work to take the Best of British edtech companies to the world and be a launchpad for the world's best education and learning organisations to base themselves and grow in the UK. (Edtech UK 2015)

Edtech UK has been established by the Education Foundation, which describes itself as 'the UK's first independent, cross sector, education think tank' and is 'focused on three priorities: education reform, technology and innovation'. Since 2011 it has led an 'edtech incubator' for new educational technology companies; worked with Facebook on a guide for educators; sought to influence policy development at a national level including running Britain's first Education

Reform Summit in partnership with the Department for Education and the Secretary of State for Education; developed a corporate partners network with Facebook, IBM, Pearson, HP, Randstad Education, Cambridge University Press, McKinsey, Skype, Sony, Google and Samsung; and delivered policy roundtables, conferences, summits, and media events around educational technology in both the UK and USA. Itself an 'incubated' project of the Education Foundation, Edtech UK was launched by Boris Johnson, then Mayor of London, with the support of the UK government departments of Business, Innovation and Skills, and of Trade and Industry, as well as by a private sector coalition of organizations from the technology sector.

Political aspirations and financial capacity, as well as technical expertise and a vision of the future of educational technology, are all combined in the activities of Edtech UK. It has powerful political support, it is modelled on financial lobbying and accelerator organizations, and it mobilizes a hybrid discourse of investment, venture capital, startup and scale-up, and economic growth. Its corporate brochure for attracting new edtech startups to London promises extraordinary benefits. It references a 'large and profitable market' for educational technology; the benefits of 'flexible procurement' regulation which allows schools autonomy in their choice of technology suppliers; proximity to global edtech companies like Pearson and Knewton and the presence of 'talent, venture capital, co-working space, government support, seed funding and events' in London; seed enterprise investment, tax breaks and 'entrepreneurial relief' for early-stage companies; plus, it claims, the incentives of 'global education technology sector spending at $67.8bn in 2015 and a global "e-learning" market worth $165bn, which is poised to reach $243.8bn by 2022' (Education Foundation 2015). It is only amid the political, financial and commercial activities of Edtech UK that the work of programmers in producing educational technologies can take place.

Edtech UK is a compelling example of how the digitization of education – through support for new edtech startup companies – relies on the financial flows that make up the lines of information in a bank account, as well as on establishing political lines of linkage, as much as on the lines of code that actually make the software work. As Lynch (2015) conceptualizes it in *The Hidden Role of Software in Education*, a new kind of 'software space' made of code, algorithms and data produced by commercial actors, programmers and analysts is nowadays working alongside both the 'economic space' of investment, funding and finance and the 'political space' of educational policymaking and governance, then exerting its influence on the 'practice space' of teaching and learning. Edtech UK is emblematic of how imaginative future visions, software, economics and politics combine and interrelate with one another to impact on the practice spaces of education. The digitization of education is not simply about the translation of educational practices into software products, but about the manifold

ways in which code comes into being, according to particular values, priorities and objectives, and in accordance with specific kinds of aspirations for the future of education.

The Digital Imagination and Materiality of Education

The examples of AltSchool, Edtech UK, Center for Data Innovation and HackingEDU we have encountered so far provide us with some sense of the imagined possibilities of datafication and digitization being associated with education. The aim of this book is neither to uncritically celebrate these developments, nor to debunk them. Instead, my intention is to consider how the twin processes of datafication and digitization are emerging from, and simultaneously reinforcing, a particular kind of reimagining of the future of education. Some sense of this reimagining is apparent from AltSchool's emphasis on personalized learning supported by data analytics platforms, and from Edtech UK's involvement in seeking to grow a future edtech market through both business and political networks. How to make sense of the work of imagination that underpins these diverse and emerging approaches?

In order to do this kind of analysis, I make use of the concept of 'sociotechnical imaginaries' from the field of science and technology studies (STS). By sociotechnical imaginaries, what is meant are 'collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understandings of forms of social life and social order attainable through, and supportive of, advances in science and technology' (Jasanoff 2015: 4). Sociotechnical imaginaries are not just science fiction fantasies: they constitute the visions and values that catalyse the design of technological projects. The dreamscapes of the future that are dreamt up in science laboratories, technical R&D departments, software companies and entrepreneurs' offices sometimes, through collective efforts, become stable and shared objectives that are used in the design and production of actual technologies and scientific innovations – developments that then incrementally produce or materialize the desired future. Through sociotechnical imaginaries, transformative scientific ideas, technological objects and social norms become fused in practice and help to sustain social arrangements or create new rearrangements in cultures, institutions and routines. Sociotechnical imaginaries are therefore the product of specifically political acts of imagination, because they act as powerful aspirational and normative visions of preferred forms of social order.

The concept of sociotechnical imaginaries has been taken up to understand the visions and values that underpin digital developments such as social media and search engines. The capacity to imagine the future is becoming a powerful constitutive element in social and political life, particularly as it infuses the technological

visions and projects of global media companies (Mager 2016). Organizations such as Google and Facebook, Apple and Amazon can be understood as dominant producers of sociotechnical imaginaries, whose aspirations are therefore becoming part of how collectively and publicly shared visions of the future are accepted, implemented and taken up in daily life. As a variation on the term 'sociotechnical imaginary', Mager (2015: 56) describes 'algorithmic imaginaries' that emerge from 'a very specific economic and innovative culture' associated with Silicon Valley technology companies, and which privilege their originators' 'techno-euphoric interpretations of Internet technologies as driving forces for economic and social progress'.

The production of such desirable imaginary futures is both social and technical, which is why they are referred to as 'sociotechnical'. That is to say, such futures are produced by particular social groups within specific social contexts, and they are also projected through the design of particular kinds of technologies – or express a view of particular futures in which those kinds of technologies are imagined to be integral, embedded parts. Unpacking sociotechnical imaginaries requires research that focuses on 'the means by which imaginaries frame and represent alternative futures, link past and future times, enable or restrict actions in space, and naturalize ways of thinking about possible worlds' (Jasanoff 2015: 24). In slightly different terms, the imagining of a 'digital future' projects a kind of 'mythology' (a set of ideas and ideals) that animates, motivates and drives forward technical development but is always much more contested and messily realized, and never as simple, straightforward or idealized as it is imagined to be (Dourish and Bell 2011).

Imaginaries in this sense act as models or diagrams to which certain actors hope to make reality conform, serving as 'distillations of practices' for the shaping of behaviours and technologies for visualizing and governing particular ways of life and forms of social order (Huxley 2007: 194). Sociotechnical imaginaries animate technical projects and social organization, and provide models for ways in which certain spaces and places might be designed and arranged. The organization of societies in this sense depends on shared imaginative resources, language and practical techniques that combine in the materiality of 'fabricated spaces' – that is, spaces that have been 'realized' in the form in which they have been imagined (Rose 1999a: 33). In other words, sociotechnical imaginaries are often enacted and materialized through linguistic and concrete practices in ways that weave the underlying vision into the fabric of society. Thus, while sociotechnical imaginaries 'can originate in the visions of single individuals or small collectives', they can gather impetus 'through blatant exercises of power or sustained acts of coalition building' to enter into 'the assemblages of materiality, meaning and morality that constitute robust forms of social life' (Jasanoff 2015: 4). Fabricated spaces, then, are the result of imaginaries that have been realized and materialized through particular technical, discursive and practical acts.

We can understand new educational projects and places such as AltSchool as the fabricated material product of a specific sociotechnical imaginary of education. It has been brought into existence as a new fabricated space of education through discursive and material means as ways of realizing a future that is seen by its advocates and sponsors as desirable and possible to attain. In other words, AltSchool itself acts as an imaginary model for the future spaces of schooling that it is seeking to fabricate in reality through operationalizing its technical platforms, and which it is supporting discursively through reference to specific kinds of progressive educational thinking. Moreover, we can think of AltSchool as an extension of Silicon Valley, translating its particular culture and spaces of innovation to the education sector. AltSchool represents the sociotechnical imaginary of Silicon Valley relocated to the materiality of the classroom. Given AltSchool's aspirations to scale its model to other sites, we can appreciate how AltSchool functions as the material product of a sociotechnical imaginary which defines how education in the future might be, could be, or perhaps even *should* be, and that might shape and delimit the everyday practices of all those who inhabit it. In this sense, the current sociotechnical imaginaries and mythologies of education, in which digitization and datafication will play a significant role, are already becoming the lived reality of education – with all of the mess and potential contestation that entails – and need to be critically examined for the material effects they might exert.

Researching Digitization and Datafication In Education

If imaginary spaces become material zones to inhabit, they can therefore exert real consequences on those who experience them. To tease open the material consequences of emerging sociotechnical imaginaries of education, it is important to look closely at the software that will make such spaces operational. Researching the digitization and datafication of education therefore requires some novel methodological and conceptual approaches. Although the science, technology and society (STS) concept of sociotechnical imaginaries can help to understand the future visions that are animating and catalysing recent and ongoing technical development, we also need methods and concepts to grasp their (actual or potential) material consequences and effects. The emerging field of digital sociology has begun to address how digital technologies, software and data are being embedded into all kinds of social and cultural activities, institutions, relations and processes (Orton-Johnson and Prior 2013):

> For some theorists, the very idea of 'culture' or 'society' cannot now be fully understood without the recognition that computer software and hardware devices not only underpin but actively constitute selfhood, embodiment, social life, social relations and social institutions. (Lupton 2015a: 2)

For digital sociologists, digitization has important implications for our ways of knowing, studying and understanding the social world, which demand interdisciplinary approaches drawing from a longer history of internet studies, media and cultural studies, science and technology studies, surveillance studies and computational social science (Daniels et al. 2016; Halford et al. 2013).

Digital sociology, then, confronts the ways in which 'new digital media, the data they produce and the actors involved in the collection, interpretation and analysis of these data' now increasingly structure and shape the social world (Lupton 2015a: 17–18). It seeks to understand, for example, how people's everyday lives are increasingly mediated through routine digital transactions with governments, commercial organizations and public institutions; how space is experienced through mobile devices; how social media has become part of social networks; and how we learn about the world through new digital media forms. Many of the central preoccupations of sociologists, such as identity, power relations and inequalities, social networks, structures and social institutions, now need to be considered from the perspective of the ongoing digitization and datafication of many aspects of society.

'Software studies' has emerged as an interdisciplinary orientation to the study of software, and includes research from the arts, philosophy, humanities, geography, cultural studies and the social sciences. Studies of software tend to share two key emphases. They focus on the software, programs and social cultures that produce effects in social life from a critical social scientific and cultural perspective, and on the social and material work that contributes to its production. Software studies seek to engage with the 'stuff of software' and:

> to see behind the screen, through the many layers of software, logic, visualization, and ordering, right down to the electrons bugging out in the microcircuitry, and on, into the political, cultural and conceptual formations of their software, and out again, down the wires into the world, where software migrates into and modifies everything it touches. (Fuller 2008: 1)

This is clearly a tall methodological order, requiring expertise in the technicalities of software, the political and cultural processes involved in its production, and the social consequences that occur as it then spreads into highly diverse practices of work, leisure, politics, culture, economics, social relations and so on.

In order to establish a set of methodological parameters for such research, Kitchin and Dodge (2011: 246) have usefully defined a 'manifesto for software studies':

> Rather than focus purely on the technical, it fuses the technical with the philosophical to raise questions about what software is, how it comes to be, … how it does work in the world, how the world does work on it, why it makes a difference to everyday life, the ethics of its work, and its supporting discourses. Software studies then tries to prise open the black boxes of algorithms, executable files, [database] structures, and information protocols to understand software as a new media that augments and automates society.

Their manifesto particularly highlights the need for critical research on the ways in which code emerges, how it performs, and how it seduces and disciplines. In terms of how code emerges, they urge for greater attention to the knowledge, practices, materials and marketplaces that are involved in the production of code, and the political, economic and cultural contexts that frame its production. They suggest performing detailed ethnographic studies of how developers produce code, and the life of software projects, to understand how software is created and how it is put to work in specific contexts.

Kitchin and Dodge then suggest that software studies might attend to the ways in which code performs. By this they mean analysing in detail 'the contextual ways in which code reshapes practices with respect to industry, transportation, consumption, governance, education, entertainment and health', as well as 'knowledge production, creative practice, and processes of innovation', and studying how code 'makes a difference' to those spaces and contexts through imbuing them with the capacity to do new types of work (Kitchin and Dodge 2011: 249). They also argue that code seduces and disciplines, largely because it offers people real benefits in terms of convenience, efficiency, productivity and creativity, whilst also enforcing more pervasive forms of surveillance and management. In particular, Kitchin and Dodge note how software is supported by powerful and consistent discourses, such as those of safety, security, empowerment, productivity, reliability, economic advantage, which persuade people to willingly and voluntarily embrace it. As such, software and code are amenable to forms of documentary and discourse analysis.

'Critical data studies' is another emerging body of interdisciplinary research that engages with the datafication of many aspects of society. A special issue on the topic of critical data studies introduced the field as a 'formal attempt at naming the types of research that interrogate all forms of potentially depoliticized data science and to track the ways in which data are generated, curated, and how they permeate and exert power on all manner of forms of life' (Iliadis and Russo 2016: 2). Iliadis and Russo (2016: 5) further highlight the identification of social data problems and the design of critical frameworks for addressing them. As a set of approaches to the critical examination of various forms of digital data – including big data, open data and data infrastructures – as well as the diverse practices of data science as a social, professional and technical discipline, critical data studies has found purchase with geographers, sociologists, philosophers and researchers of education.

In one of the first publications detailing critical data studies, the geographers Dalton and Thatcher (2014) set out seven defining commitments: (1) situate data regimes in temporal and spatial context; (2) reveal data as inherently political and expose whose interests they serve; (3) unpack the complex, non-deterministic relationship between data and society; (4) illustrate the ways in which data are never raw but always intentionally generated; (5) expose the fallacies that data can speak for themselves and that exhaustive big data will

replace smaller-scale sampled data; (6) explore how new data regimes can be used in socially progressive ways; and (7) examine how academia engages with new data regimes and the opportunities of such engagement.

In another article outlining concepts and methods for critical data studies, Kitchin and Lauriault (2014) seek to provoke researchers to unpack the complex 'assemblages' that produce, circulate, share/sell and utilize data in diverse ways. Data assemblages, as they define them, consist of technical systems of data collection, processing and analysis, but also the diverse social, economic, cultural and political apparatuses that frame how they work. In this broad sense, a data assemblage includes: (1) particular modes of thinking, theories and ideologies; (2) forms of knowledge such as manuals and textbooks; (3) financial aspects such as business models, investment and philanthropy; (4) the political economy of government policy; (5) the materiality of computers, networks, databases and analytics software packages; (6) specific skilled practices, techniques and behaviours of data scientists; (7) organizations and institutions that collect, broker or use data; (8) particular sites, locations and spaces; and (9) marketplaces for data, its derivative products, its analysts and its software.

Approaching critical data studies in terms of sociotechnical data assemblages is productive for research into the production and use of educational data. This book provides a series of explorations of big data as it is entering into the complexities of education and reworking teaching, learning, assessment, governance and educational research itself. For the field of education research, big data is a new and emerging phenomenon about which there remains limited knowledge (Beneito-Montagut 2017). In the following chapters, I combine the focus on sociotechnical imaginaries with digital sociology, software studies and critical data studies approaches as a methodological strategy to perform a series of critical analyses of the ways in which assemblages involving software code, algorithms and digital data are making a difference in education.

This is not to suggest that existing approaches to educational research, description and explanation are irrelevant. Rather, part of my aim is to demonstrate that educational research can be productively extended by engaging with software and data from a critical perspective. Studies of educational policy, for example, have already begun to engage with the software packages and data infrastructures that enable policy information to be collected, and that also allow policies to penetrate into institutional practices. In the following chapters I seek to understand how some of the software technologies penetrating education today have come into existence and inquire into the imaginaries that animate them; to explore the forms of expertise and knowledge they work in relation with; to examine how they are being put to work in specific contexts and spaces and how they are shaping particular practices; and to explore how they are promoted and supported by certain discourses emanating from diverse public, private and philanthropic sectors.

Learning Machines

By working with concepts of sociotechnical imaginaries and critical approaches to software and data, I aim to show how powerful future visions are fast being turned into the ordinary artefacts that are enabling digitization and datafication in education. A useful term to capture these artefacts of educational digitization and datafication is 'learning machines'. This is a term I borrow from Michel Foucault. In his highly influential work on regimes of discipline, Foucault (1991) traced some of the ways in which schools function to supervise and discipline pupils, particularly through techniques like timetabling, sitting them in rows in classrooms, and organizing them in ranks according to age, performance, behaviour, knowledge and ability. Together, Foucault (1991: 147) argued, these techniques 'made the educational space function like a learning machine, but also as a machine for supervising, hierarchizing, rewarding … according to the pupils' progress, worth, character, application, cleanliness and parents' fortune.' He detailed how classrooms functioned by placing pupils in categories, classifications and rankings based on constant assessments of their qualities, age, development, performance and behaviour. Through techniques of ordering and ranking pupils according to diverse categories, Foucault argued, 'the classroom would form a single great table, with many different entries', and he noted that classrooms are 'mixed spaces' – 'real' insofar as they consist of buildings, rooms and furniture, but 'also ideal, because they are projected over this arrangement of characterizations, assessments, hierarchies' (1991: 148).

The categorization and tabularization of educational institutions, spaces, processes and individuals is perhaps the ideal aim – or dominant imaginary – of big data in education. In this sense, what Foucault designated learning machines takes on new resonance in the era of big educational data. The learning machines being imagined and built today consist of computational technologies that can capture and process data about learning; that can intervene in learning practices, processes and institutions; that can 'learn' from the data they process; and that can be understood as techniques of power, instruments for the control of activity, behaviours and bodies, and processes of knowledge generation. They are smart learning machines, the material and operational form of the sociotechnical imaginary of big data in education. Through big data, schools, colleges, universities and other informal learning contexts are becoming 'machine[s] for learning, in which each pupil, each level and each moment, if correctly combined', are becoming 'permanently utilized in the general process of teaching' by a 'precise system of command' which operates 'according to a more or less artificial, prearranged code' (Foucault 1991: 165–6). The smart learning machines associated with digitization and datafication in education are the product of lines of code, in the technical sense, that also enforce particular codes of conduct. Digital software allows institutions, practices and people to be

constantly observed and recorded as data; those data can then be utilized by learning machines to generate insights, produce 'actionable' intelligence, or even prescribe recommendations for active intervention. The ideal sociotechnical imaginary of big data in education is now being materialized and operationalized through smart learning machines, made of software code and data, which might inhabit real educational spaces.