

Mertala, P. (accepted). Data (il)literacy education as a hidden curriculum of the datafication of education. *Journal of Media Literacy Education*.

Abstract

This position paper uses the concept of “hidden curriculum” as a heuristic device to analyze everyday data-related practices in formal education. Grounded in a careful reading of the theoretical literature, this paper argues that the everyday data-related practices of contemporary education can be approached as functional forms of data literacy education: deeds with unintentional educational consequences for students’ relationships with data and datafication. More precisely, this paper suggests that everyday data-related practices represent data as cognitive authority and naturalize the routines of all-pervading data collection. These routines lead to what is here referred to as “data (il)literacy”—an uncritical, one-dimensional understanding of data and datafication. Since functional data (il)literacy education takes place subconsciously, it can be conceptualized as a form of hidden curriculum, an idea that refers to lessons taught and learned but not consciously intended to be so.

INTRODUCTION

It is Tuesday morning, and 8-year-old Milla enters school. First, she registers her attendance by having her personal near-field communication (NFC) tag read by a monitor in the entrance of her home classroom. Her first class is math. Milla and her classmates use a ViLLE learning environment containing real-time learning analytics. In her second class, the children read out loud to a mobile application called “Luppakorva,” which records and analyzes their reading. As Milla’s school has a bring-your-own-device policy, children are using their personal phones. After lunch, Milla’s class has two hours of physical education, during which the teacher gives each child an activity wristband with an integrated heart rate monitor. The children’s heart rates are displayed on the wall of the gym, and the children receive points for the durations in which they operate at the target heart rate level. The software creates a report on each child’s performance available for parents and children to observe after school.

The narrative above is constructed from various sources (Ervasti et al., 2010; Song, 2014; Williamson, 2017b; Kurvinen et al., 2019; Rytönen, 2019) that have described the quotidian digital data-related practices in schools and all the technologies mentioned—NFC tags, the ViLLE platform, the Luppakorva app, and activity wristbands—many of which profile students based on their data. The purpose of this narrative is to concretize the ways digital datafication has

woven itself into the everyday fabric of contemporary education, to paraphrase Weiser's (1991) famous notion. Indeed, the "datafication" of education, as it has been called, has been identified as "one of the defining issues of contemporary education" (Selwyn, 2018, p. 734).

The key argument of this position paper is that everyday data-related practices, such as the ones mentioned above, can be approached as functional forms of data literacy education: deeds with unintentional educational consequences (Siljander, 2002) concerning students' relationships with data and datafication. More precisely, this paper argues that everyday data-related practices in education represent data as cognitive authority to students (Wilson, 1983) and naturalize the routines of all-pervading data collection (Couldry & Yu, 2018). These routines lead to what is here referred to as "data (il)literacy"—an uncritical, one-dimensional understanding of data and datafication. Since functional data (il)literacy education takes place on a subconscious level, it can be conceptualized as a form of "hidden curriculum," a term that refers to lessons taught and learned but not consciously intended to be so (Kentli, 2009).

This article is structured as follows. First, an account of how data and datafication are understood in this paper is provided. The context of formal education is then brought into focus by discussing how datafication relates to digitalization (Selwyn, 2019a), learnification (Biesta, 2012), and accountability (Biesta, 2004), which are other meaningful determinants of contemporary schooling. An overview of data literacy and data literacy education in educational research and praxis is then given. The remaining sections are reserved for presenting the different forms of the hidden curriculum of datafication and their pedagogical outcomes. In each of these sections, examples and cases from the research literature and public accounts (e.g., news pieces, company websites) are provided to concretize the phenomenon under discussion. The examples cover different national contexts (e.g. Finland, China, USA), various stages of education (e.g., early childhood education, primary education, higher education), and a wide range of technologies (e.g. learning analytics, facial recognition technologies) to illustrate the pervasiveness of the datafication of education.

PUTTING THE DATAFICATION OF EDUCATION IN CONTEXT

Data and datafication

Typically, the term "data" is accepted to mean "numbers" or "quantified evidence" (Bowler et al., 2017)—the raw material produced by abstracting and reducing the world into representative forms (Kitchin, 2014). Such definitions, however, are rather technical by nature, and various authors (e.g., Kitchin, 2014; Williamson, 2017) have advocated for adopting a more socio-technical perspective on data, underlining that data are never raw but always intentionally generated. Put differently, the questions should be what data are collected, how

the data are collected, what the data are believed to represent, and for what the data are used—questions that are determined by social agents with varying intentions, needs, and desires, entailing that data are never purely neutral or objective.

Datafication, then, refers to the process whereby most of our everyday practices, both online and offline—including aspects of the world not previously datafied and measured, such as social relations and emotions (Mascheroni, 2018)—are converted “into online quantified data, thus allowing . . . real-time tracking and predictive analysis” (Van Dijck, 2014, p. 198). As pointed out by several authors (e.g., Breiter & Hepp, 2018; Mascheroni, 2018; Sadowski, 2019), datafication can be considered a defining phenomenon of our contemporary mediated lifeworld. Various examples support this claim. First of all, virtually all our technology-mediated actions generate digital data: All photos and videos taken via smartphones or action cameras, such as GoPro, contain metadata (e.g., locations, dates, and times) that are mostly invisible to the user but ready to be “harvested by the company providing the service” as well as “processed through algorithms to detect people, places, brands, and even emotions” (Slotte Dufva, n.d.). In addition, almost all websites contain trackers that collect and correlate data about the Internet activities of particular users, computers, and devices across time (Center for Democracy and Technology, 2011). The tracking is done by (often commercial) sites themselves or by third-party trackers, such as Google Analytics (Bailey et al., 2019). As these examples illustrate, many data are produced as an unintended side effect of our technology use and online activities, and only a limited group of users are therefore aware of the scope of datafication (Breiter & Hepp, 2018; see also Bowler et al., 2017; Pangrazio & Selwyn, 2018).

These data are often collected and further used to understand, predict, and influence our decisions. Some relevant examples are the recommendation systems that track the data of many people to very accurately predict their interests based on people with similar interests and content similarity (Valtonen et al., 2019). To give an idea of the effectiveness of (high-quality) recommendation systems, an estimated 80% or more of the television shows and movies people watch on Netflix are discovered by through the platform’s recommendation system (Blattman, 2018). Put differently, when one chooses what to watch on Netflix, one essentially chooses from many data-informed decisions made by an algorithm. Such data-driven algorithmic services have been conceptualized as persuasive technologies, interactive computing systems designed to change attitudes or behaviors (Fogg, 2003).

Another set of timely examples are self-tracking devices, such as activity wristbands, sport/smart watches, and smart rings like Oura, as well as various mobile applications (apps) (e.g., Sport Tracker, My Fitness Pal, etc.). These data-collecting devices are highly popular for tracking health and physical

performance and can be paired with many apps and websites that support user-led data collection and allow users to interpret and visualize their own health data (Williamson, 2017b). The breadth of acceptance these devices and apps enjoy is perhaps illustrated by looking at the numbers: More than 100,000 health apps are available (Lupton & Jutel, 2015), some of which are highly popular. At the time of writing, the Adidas running app (Adidas Running app, n.d.) has over 50,000,000 downloads in Google Play, and the activity tracker app for FitBit has more than 10,000,000 downloads (Fitbit, n.d.). In 2017, the global unit shipment of sport watches reached about 18.6 million units (Statista, 2020). Self-tracking also differs from many other forms of digital surveillance in that the data are collected at the users' own discretions to optimize certain aspects of their lives, including health.

Datafication is not limited to adults. Children, too, are "objects of [a] multitude of monitoring devices that generate detailed data about them" (Lupton & Williamson, 2017, p. 780). The datafication of childhood takes place in various forms, as the numerous downloads of pregnancy and parenting apps, the increasing sales of wearable devices aimed at babies and children, and the growing market of Internet-connected toys all show (Mascheroni, 2018). As soon as children own smartphones, the amount of data collected from them increases rapidly, as mobile phone ownership makes the Internet much more available (Merikivi et al., 2016) and intensifies the collection of data. When children enter the formal education system, these forms of data-based surveillance (known as "dataveillance") and datafication are complemented by many others (Lupton & Williamson, 2017). The introductory section presented some of these forms, and other examples are discussed below.

Datafication, learnification, digitalization, and accountability in education

Students' attendance and performance have been monitored throughout the history of formal education via checklists and the systematic (manual) recording and tracking of exam scores (Selwyn, 2018). However, as in every other sector, the advent of automatically collected and analyzed (big) data in education has exploded both the breadth and depth of data collection to unprecedented levels. That said, it is important to acknowledge that datafication is not an isolated phenomenon in the context of education; it intertwines with various other phenomena, the most profound being digitalization, learnification, and accountability (see also Bradbury & Roberts-Holmes, 2017; Williamson, 2017a). The rapid development of digital technologies has enabled the development of (big) data-driven practices in schools and other educational settings, something this paper refers to as the "interconnectedness" of digitalization and data. As Selwyn (2019a, p. 79) noted,

The ubiquity of personal digital devices (not least smartphones, tablets and laptops) ensures that most schools and universities operate in a state

of “one-to-one” access where every student and teacher has access to at least one personal device at any time. This allows educational institutions to operate through large-scale platforms, such as the all-encompassing “learning management system.” . . . Crucially, all these technologies facilitate the continuous generation and processing of large quantities of data. This data relates to most aspects of education—ranging from the individual action of students to institution-wide processes of performance.

The notion of performance-monitoring leads us to accountability, which generally bears connotations of being answerable to someone (Biesta, 2004). In the educational context, accountability—as a transnational policy trend (Lingard et al., 2013)—refers to measurement and statistical analyses to evaluate educational outcomes (Paananen, 2017) where data are used for comparisons between schools, and students as well as within individual subjects to compare past performance to present (Bradbury & Roberts-Holmes, 2017; Williamson, 2017a). For example,

[The] United Kingdom’s National Pupil Database contains detailed data on over 7 million British schoolchildren from 2002 onwards, constituting one of the largest educational datasets in the world. These linked datasets, combined with databases of information from further and higher education, enable individual pupils to be monitored throughout the educational life course. (Lupton & Williamson, 2017, p. 784)

As the idea of being answerable to (Biesta, 2004) suggest, such data has social consequences. For example, data about student performance –namely learning outcomes measured by regular standardized tests– are used as evidence to evaluate teacher and school performance (Lewis & Holloway, 2019; Stevenson, 2017) and in some cases teachers have been fired if their students perform weaker than the data-based model predicts (O’Neil, 2016).

The examples above also connote with the interrelatedness of datafication and learnification which can be traced back to the trend that questions around education tend to be reduced to questions of learning (Biesta, 2012). While learning is a complex process, the indicators and concepts of learning appear more amenable to measurable, quantifiable forms than, say, educational objectives like enabling good lives and producing good humans (Buber, 1937), which are ambiguous, undeterminable qualities. For example, it has been argued that collaboration—which is often argued to be a prerequisite of effective learning (Baker, 2015)—can be detected by comparing students’ arousal-directional agreement (Pijeira-Diaz et al., 2019); shared levels of high arousal (measured in numerical form) between two or more group members signals the existence of collaboration. Learning processes can be detected by measuring

students' electrodermal activity—namely, changes in the electrical conductivity of the skin dynamics of collaborative learning can be captured and measured from the “commonalities and interdependence in the degree of physiological activation from the sympathetic nervous system (i.e., sympathetic arousal) of group members” (Pijera-Diaz et al., 2019, p. 188).

To conclude, datafication, digitalization, learnification, and accountability form an assemblage in which the parts constantly interact with one another. Perhaps the most powerful example to concretize their interrelated nature are learning analytics. As Selwyn (2019b, p. 14) noted, “One of the core tenets of learning analytics is that data (in particular, data derived from digital technologies in educational contexts) can 1) be used to model learning processes that have taken place; and 2) thereby provide a basis for making decisions regarding future learning.” This “core tenet” neatly illustrates how data are captured and analyzed via digital technologies (digitalization) to produce evidence about students' learning (learnification) as the basis of decision-making (accountability).

DATA LITERACY AND DATA LITERACY EDUCATION

Data, literacy, and education are all ambiguous, multidimensional concepts. This complexity by no means declines when they are combined. Data literacy, for example, is a fluid concept with no universally accepted meaning (Bowler et al., 2017; Pangrazio & Sefton-Green, 2019). One reason for this ambiguity is that discussions around data literacy take place in various scientific fields and from the perspectives of multiple empirical contexts (see Koltay, 2015; Pangrazio & Selwyn, 2019). Thus, as Koltay (2015) has noted, data literacy has no distinguished identity. Rather, it “falls into the same concept pool as multi-literacy, digital literacy, information literacy, digital media literacy and media literacy” (Markham, 2020, p. 229).

Given this ambiguity, one could validly question the need or use for the concept of “data literacy.” Pangrazio and Selwyn (2019) however, grounded the justification of data literacy as an independent concept in the growing significance of personal data. Drawing on the traditions of critical literacies and the critical strand of New Literacies studies, they argued for “a need to better support individuals to engage critically with their personal data so they have a sense of understanding, control and agency within the data assemblage” (Pangrazio & Selwyn, 2019, p. 427)—in other words, data literacy. They are not alone in their view; many scholars have recognized the importance of data literacy as a transversal competence that all citizens should possess in increasingly data-driven societies (e.g., Wolff et al., 2016; Pangrazio & Sefton-Green, 2019) and beyond (e.g., Spin, 2017; Bhatia, 2018).

One concrete example of the growing interest in data literacy is that the need for systematic data literacy education is regularly invoked in scholarly and public

discussions (e.g., Wolff et al., 2016; Spina, 2017; Bhatia, 2018; Gebre, 2018; Schuff, 2018; Pangrazio & Sefton-Green, 2019). In these discussions, data literacy education has typically been defined and presented as teachers' intentional pedagogical interventions to teach students how to read and use data effectively (Wolff et al., 2019). Pangrazio and Sefton-Green (2019, pp. 8–9) referred to these approaches as “formal data literacy pedagogies” that often “prioritise the positive utility of data, showing students and teachers how they can do better research, enact social change or improve decision-making.” In practice, formal data literacy education employs both large-scale external data sets and small-scale data sets collected by students, and it is typically organized as an independent subject in the form of inquiry-based projects (Gebre, 2018; Wolff et al., 2019).

What appears to be missing from data literacy education are approaches that connect data literacy to students' everyday digital lifeworlds. Such approaches seem needed, as many students' understanding of everyday datafication is limited (Bowler et al., 2017; Gebre, 2018; Pangrazio & Selwyn, 2018). Data is often conceptualized in terms of experiments or survey data (Gebre, 2018), not as data generated automatically by common online activities (Bowler et al., 2017; Gebre, 2018). Accordingly, many students fail to recognize the collection of geo-locational data, for instance, as a form of datafication (Bowler et al., 2017; Pangrazio & Selwyn, 2018). Research has suggested that formal data literacy education does not contribute to improving student understanding of the aforementioned issues (Bowler et al., 2017; Gebre, 2018), implying the need for more contextual approaches, which is supported by an empirical study by Pangrazio and Selwyn (2018). They taught data literacy to 13–17-year-old students using an app that aggregated students' personal data and demonstrated to each participant how their data might be recirculated and reused by various third parties. According to Pangrazio and Selwyn (2018), this method allowed the students to become more conscious of geo-locational tracking and the precision with which it could trace their movements. Building partly on these experiences, Pangrazio and Selwyn (2019) called for data literacy education that moves beyond the technical, value-free connotations of data to include the political economy of the digital platforms of datafication. This “critical approach to data education ... seeks to raise consciousness of the social injustices associated with datafication and help students to question and challenge dominant ideologies, beliefs and practices” (Pangrazio & Selwyn, 2020, p. 5).

Such a curriculum would undoubtedly be more holistic and contextualized than currently prevalent forms of data literacy education. Nevertheless, without critical reflection on the political and commercial aspects of the datafication of education, it would remain superficial at best. To a great extent, the data-related practices of contemporary education replicate those from other sectors and are at least partially driven by major technology and data companies, such as

International Business Machines (IBM) and Pearson (Williamson, 2017a). Thus, it might smack of hypocrisy to teach students about the political economy of datafication using external examples and cases while predisposing them to use such technologies and practices as part of everyday schooling. The whole idea of approaching data literacy education exclusively as formal, teacher-led lessons is based on a rather restricted understanding of education. Besides intentional pedagogical actions, the everyday practices of institutional education are pregnant with actions that have notable, though unintended, educational consequences. Siljander (2002) referred to these deeds as “functional education.” This paper thus broadens the idea of data literacy education to include both teacher-intended formal data literacy education *and* unintentional data educational that occurs through largely unexamined, quotidian data-related practices.

Drawing on the terminology of curriculum studies, teacher-intended data literacy education can be defined as the “official” (Giroux & Penna, 1979) or “formal” (Portelli, 1993) curriculum of data literacy education. These school/classroom level meso/micro curricula are typically guided by macro curricula provided by the state, which, in turn, are influenced by supranational agents, such as the Organization of Economic Co-operation and Development and the European Union (Erstad & Voogt, 2018; see also Palsa & Mertala, 2019). Everyday data-related practices, on the other hand, constitute a “hidden curriculum,” which refers to lessons taught and learned but which are not openly or consciously intended to be such (Kentli, 2009). These lessons are not guided by the macro and/or meso curricula; instead, in the case of datafication, the hidden curriculum taught in schools replicates and legitimates—to a notable extent—the logics and practices of commercial agents, such as technology companies. In summary, teachers can simultaneously implement intended and hidden curricula in relation to data literacy.

THE HIDDEN CURRICULUM OF DATAFICATION OF EDUCATION

The education sector is one of the most noticeable domains affected by datafication, because it transforms not only the ways in which teaching and learning are organized but also the ways in which future generations (will) construct reality with and through data (Jarke & Breiter, 2019, p. 1).

The quotation above neatly captures how datafication transforms education and its outcomes. Put differently, datafication not only shapes the ways education is provided but also contributes to shaping students’ relationships to and understandings of data and datafication. No single form of hidden curriculum is straightforwardly deterministic or all-encompassing, of course. First of all, not all students are alike; while many struggle to understand the breadth and variety of datafication (Bowler et al., 2017, Gebre, 2018; Pangrazio & Selwyn, 2018a),

some possess more conscious, agentic stances (Goodyear et al., 2019). All the aforementioned applies to teachers as well; it would be an oversimplification to claim that all teachers adopt uncritical attitudes toward the datafication of education. Nevertheless, the more space and power given for data-collecting and data-processing technologies in schools, the bigger the effect they will have on teachers' and students' choices and actions (Selwyn, 2019a).

Historically, hidden curricula have taken various forms (Kentli, 2009). This applies to datafication as well. Concerning data literacy, current formal data literacy pedagogies that introduce data to students as external datasets or self-collected research project data (see Gebre, 2018; Wolff et al., 2019) already contain the hidden, unintentional lesson that data is limited to these conceptualizations. While most research on hidden curricula has concentrated on human interaction, the materials and resources used in classrooms—including data-generating and processing devices and software—may also carry and teach such hidden messages (Edwards, 2014). The following sections discuss the forms and content of the hidden curriculum of datafication in more detail. The focus is on two partially overlapping themes: representing data as cognitive authority and the naturalization of all-pervading data collection. These sections also discuss the kinds of data (il)literacy these practices and routines produce.

Representation of data as cognitive authority

Cognitive authority, as defined by Wilson (1983), refers to an information source—human or non-human—that people deem credible and legitimate. The term is useful in the context of datafication, as people tend not to treat data as “proxies” or “indicators” but as direct measures (Selwyn, 2019b, p. 12). An illustrative example of this straightforward logic is the previously discussed data-based (teacher) accountability where data about students' learning outcomes are used as direct measure of teacher performance (see Lewis & Holloway, 2019; O'Neil, 2016). This phenomenon is at least partially due to the quality of discursive practices around datafication, data typically being presented as accurate and unmistakable, making them “undisputed authorit[ies]” (Špiranec et al., 2019) and a “superior form of evidence” (Battista & Conte, 2016, p. 147) for decision-making. A glance at the ways datafication is advertised to the educational sector illustrates that such views also exist in the context of formal education. The use of learning analytics is promoted to enable personalized learning, which is typically argued to provide two kinds of benefits. First, as put by Dural and Gros (2014, p. 383) they are “powerful tool[s] for helping students reflect on their learning activity and, therefore, gain knowledge about their learning processes. This is especially important, since self-knowledge can be considered as a key metacognitive skill.” This argument echoes the view of data as a direct measure (Selwyn, 2019b), as it states that datafication (in the form of learning analytics) provides knowledge about the

learning processes instead of information or data that, unlike knowledge, convey the need for critical assessment, evaluation, and interpretation from the reader. Second, the use of personalized learning analytics is argued to be more effective than traditional classroom teaching, as it is impossible for teachers to perfectly differentiate instructions and exercises to meet the diverse needs of students (e.g., Ebner & Schön, 2013; Kurvinen et al., 2019). Another example is the Finnish sport technology company Polar, who endorsed their educational products with similar discursive devices by stating that “with reports from Polar products, physical education teachers can show how well students have developed, for example, for budget applications or for parents of students” (Polar, n.d.)

To summarize the key messages of the extracts and examples above, data are interpreted as accurate, objective, and valuable by those who decide budgets in the educational sector. The statement about data being a direct measure of students’ development is also an illustrative example of the intertwining relationship of datafication and accountability. These messages appear to be accepted by education providers. For example, the Finnish private kindergarten chain Touhula rationalizes the use of Polar Active tracker wristbands by highlighting that the devices are

specifically designed to measure the amount and intensity of children’s exercise. The activity tracker provides easy and clear data regarding the day: how much the kids have been sitting, standing, or moving around. With the aid of the measured data, tracking the quality of activities is easy (Touhula, n.d.).

The problem is the limited correspondence between the discursive and practical levels of data and datafication, as the data are mere proxies and indicators of the phenomena the data collection is claimed to capture (O’Neil, 2016). Take activity wristbands, for example. The Polar Active wristbands used in Touhula kindergartens use accelerometer technology to detect their users’ physical activities, which they measure by the movement of subjects’ hands, neglecting forms of physical activity in which hands are static (e.g., riding a bicycle or tricycle or pushing a trolley) (Chen et al., 2016). These monitors also tend to consider large, continuous arm movements as step counts while sitting and standing (Chen et al., 2016), making them rather unreliable instruments to measure physical activity. Learning analytics also rely on proxies and indicators of the complex, situated, and multifaceted process of learning. Be they indicators of electrodermal activity from body sensors (Pijera-Diaz et al., 2019), performance data collected via instructional games (Kurvinen et al., 2019), or automated essay scoring (Selwyn, 2019a), each of these sources represent different technology clusters and reflect different perspectives on the social relations of knowledge and learning (Cope & Kalantzis, 2015). Instead of analyzing learning *per se*, they analyze *proxies* from discrete factors that have been identified as meaningful for learning.

As the aforementioned discursive examples show, these limitations are seldom addressed by the proponents of learning analytics (or proponents of datafication of education in general) or reflected and reproduced in the ways data are used and represented in everyday classroom situations. Concerning data (il)literacy, presenting and treating data as undisputed cognitive authority may lead students to overestimate the accuracy of data and to build excessive trust in the reliability of analyses and reports produced by devices and software. The unique nature of the student–teacher relationship intensifies this process: For students, teachers are cognitive authorities whose knowledge and actions are typically deemed legitimizing (Raviv et al., 2003; Esmaeli et al., 2018; Wang et al., 2019). Thus, if teachers present data as a “superior form of evidence” (Battista & Conte, 2017, p. 147) to students, they are (likely unintentionally) emphasizing the message by being cognitive authorities themselves.

Viewing data as cognitive authority relates to the concern that the use of data-driven technologies can reduce students’ capacities for agentic decision-making (Williamson, 2017a; Selwyn, 2019b). Williamson (2017a, p. 120), for instance, called data-driven learning analytics “decisional interference” that,

rather than engaging students in their right to involvement in decisions about important matters that affect their own lives, . . . appear to distribute decision-making to automated, proprietary systems where students have little opportunity for involvement in the handling or use of their own data.

Selwyn (2019b, pp. 12–13) discussed the same phenomenon: “While learning analytics are often framed in terms of supporting human decision-making, most often these technologies are to direct (if not determine) human decision-making.” These examples resemble the recommendation systems used by Netflix and many others discussed earlier. While choosing what to watch on a Friday night may not count as “decisions about important matters” (Williamson, 2017a, p. 120), the increasing externalization of decision-making to persuasive technologies may diminish subjects’ agency. Interestingly, persuasiveness appears to be something that students expect from datafied educational practices. Many (higher education) students in a study by Schumacher and Ifenthaler (2018) commented that learning analytics should actively contribute to regulating and shaping their behavior and actions. Some even wished that learning analytics could access their personal calendars to provide learning recommendations matching their schedules, a notion that serves as bridge to the next theme: the naturalization of all-pervading data collection.

Naturalization of all-pervading data collection

The second feature of the hidden curriculum of datafication is the naturalization of all-pervading data collection. The more datafied a schools’ practices are, the more natural and acceptable datafication appears to its students. As Couldry and

Yu (2018) pointed out, the naturalization of datafication and surveillance through discourses and routinized practices frame surveillance as a natural part of the world we inhabit and data as neutral means of achieving benefits and empowerment (see also Mashceroni, 2018). While Couldry and Yu (2018) did not make this claim in the context of formal education or hidden curricula, their ideas resonate here as well.

Take learning analytics, for example. According to Ifenthaler and Schumacher (2016), learning analytics systems require vast arrays of data to produce their expected adaptive, personalized information. These data include personal information, including online behavior outside the learning management system, as “Such data includes much potential for understanding and optimizing learning processes” (Ifenthaler & Schumacher, 2016, p. 933). While comments like these may seem like concessions to the idea that “Classrooms are not closed, computable systems based upon controllable variables that can be monitored and manipulated” (Selwyn, 2019a, p. 91), they are also arguments that maximizing the benefits of learning analytics depends on (or requires) a willingness to share as much data as needed.

Another example of the naturalization of all-pervading data collection is the growing interest in the use of facial recognition technology in schools. According to Andrejevic and Selwyn (2019), there are three drivers of this movement: security-based surveilling in schools and campuses, monitoring student attendance, and using facial detection techniques as indicators of student engagement and learning. Whatever the motivation, facial recognition technology collects enormous amounts of identifying data from students. For example, the facial recognition system used in Hangzhou No. 11 High School in China

scans classrooms every 30 seconds and records students’ facial expressions, categorizing them into happy, angry, fearful, confused, or upset. The system also records student actions such as writing, reading, raising a hand, and sleeping at a desk. (Chan, 2018, n.p.)

This level of scanning frequency produces 120 data points for each student every hour. This equals around 1,000 data points per day, which totals 200,000 data points per school year. While these numbers are massive, even more impressive is the lack of effort required to collect such an amount of data. Whereas data collected via learning management systems or wearable tracking devices require some kind of active input from the student, facial recognition systems collect the data silently, invisibly, and independently, and thus are an illustrative example of what Weiser (1991) meant by “disappearing technologies.”

While the use of learning analytics, wearable tracking devices, and facial recognition are forms of intended datafication, some data collection in schools

happens unintentionally. In November 2019, the Finnish National Broadcasting Company published an online article (Rytkönen, 2019) about a third-grade student who brought home documents that introduced a selection of apps to be installed on the student's mobile phone, as the school had a bring-your-own-device policy. One of the applications used by the school used the phone's microphone, recording the child's speech and home sounds. The app also reserved the right to use the information it collected for commercial purposes and to pass it on (Rytkönen, 2019). This is not an isolated case; similar incidents have been reported all over the world (e.g., Cook, 2018), and they serve as examples of how the political drive to digitize education has, metaphorically speaking, opened the classroom doors to commercial agents (see also Paakkari, 2020).

There appears to be no or little negotiation between educational administrations, students, and families around datafication policies in the educational sector. For instance, the father of the Finnish third-grade student was not asked for permission to install the apps. Instead, he was merely informed, "Hi, we're beginning to use this [app at school]" (Rytkönen, 2019)—that is, please install it on your child's phone. Schools also introduce facial recognition systems without consulting students or parents. In an interview with the *Washington Post*, Jim Shultz, the father of a 15-year-old student at a high school in upstate New York, commented that

We've [parents and students] gotten no answers to all these questions: Under what conditions can a kid's face be put into the system? Does the district need parental consent? Who can do a facial recognition search? (Harwell, 2018, n.p.)

Once again, instead of problematizing the logic and routines of datafication and dataveillance, schools have followed the same principles as software providers. If one wishes to use a certain app or service, one must comply with the data collection policies of the software provider. Likewise, if one wishes to go to school or send a child to school, one must comply with the surveillance and datafication policies and practices of that school. There are no gray areas or room for negotiation. With top-down decisions and practices like these, schools contribute to naturalizing and normalizing all-pervading data collection and the culture of constant surveillance of students. Indeed, based on media reports, many students immediately accept the new protocols and consider the surveillance systems "cool" (Alba, 2020, n.p.). By doing so, schools diminish students' possibilities for control and agency within the data assemblage (Pangrazio & Selwyn, 2019) in school and society at large and thus contribute to a form of data (il)literacy by which the students consider themselves mere passive drifters in an increasingly datafying world.

CONCLUDING REMARKS

Datafication has been called the defining phenomenon of our contemporary mediated lifeworld (Breiter & Hepp, 2018), including the educational sector (Williamson, 2017a; Bradbury & Roberts-Holmes, 2018; Jarke & Breiter, 2019). On the level of everyday praxis, the datafication of education takes the form of the increasing and intensifying use of learning analytics (Kurvinen et al., 2019), automatic surveillance systems (Andrejevic & Selwyn, 2019), and wearable tracking devices (Williamson, 2017b), to mention just a few examples.

This position paper used the concept of “hidden curriculum” as a heuristic device to analyze everyday data-related practices in formal education. Grounded in a review of research publications and public accounts of the datafication of education, this paper suggests the existence of two intertwined forms of hidden curricula. The first form, a representation of data as cognitive authority, entails that data are problematically introduced to students, not as imperfect proxies and indicators, but as direct measurements. As an unintended pedagogical outcome, students learn to overestimate the accuracy of data and build excessive trust in datafied systems. The second form, the naturalization of all-pervading data collection, implies that the more datafied a school’s practices are, the more natural and acceptable datafication and dataveillance appear to its students, which diminishes their agency. Bringing datafication and dataveillance into schools via top-down organized reforms fails to properly consult students or their parents.

While the arguments presented in this paper are grounded in a careful reading of the theoretical literature and reports of current data-related practices in formal education, they are inevitably speculative and hypothetical. Nevertheless, by suggesting that data literacy education transcends formal data literacy pedagogies, the paper provides novel, useful theoretical lenses and conceptual tools for application in future empirical research to achieve a more holistic and comprehensive understanding of datafication and its consequences in the educational sector. The two forms of hidden curricula discussed in the present paper provide theory-informed starting points for such analyses to complement the work of others (e.g., Pangrazio & Sefton-Green, 2019; Pangrazio & Selwyn, 2019, 2020) by using, for example, ethnographic methods.

Besides research, the ideas presented in this paper are meaningful for initial and continuing teacher education. While numerous publications have provided guidelines for teachers’ data literacy development (e.g., Cowie & Cooper, 2017; Mandinach & Gummer, 2013; Reeves & Honig, 2015; Schildkamp, Poortman, & Handelzalts, 2016), the take on data literacy has been restricted to training teachers to use data more efficiently as basis for decision-making and student assessment. In order to avoid the scenarios discussed in this paper, initial and continuing teacher education should include critical dimensions of data literacy as well. Training should also be tightly contextualized to the practices of everyday schooling to illustrate the risks related to implementing the hidden

curriculum of data (il)literacy. As the narrative at the beginning of the article shows, contemporary data-saturated classrooms are not short of suitable and information-rich cases.

References

- Adidas Running app (n.d.). Adidas Running app by Runtastic.
<https://play.google.com/store/apps/details?id=com.runtastic.android&hl=fi>;
- Alba, D. (2020, February). Facial recognition moves into a new front: Schools.
<https://www.nytimes.com/2020/02/06/business/facial-recognition-schools.html>
- Andrejevic, M., & Selwyn, N. (2019). Facial recognition technology in schools: Critical questions and concerns. *Learning, Media and Technology*. Advanced online publication.
<https://doi.org/10.1080/17439884.2020.1686014>
- Bailey, J., Laakso, M., & Nyman, L. (2019). Look who's tracking. *Informaatitutkimus*, 38(3–4), 20–44. <https://doi.org/10.23978/inf.87841>
- Baker, M. J. (2015). Collaboration in collaborative learning. *Interaction Studies*, 16(3), 451–473. <https://doi.org/10.1075/is.16.3.05bak>
- Battista, A., & Conte, J. (2016). Teaching with data: Visualization and information as a critical process. In N. Pagowsky & K. McElroy (Eds.), *Critical library pedagogy handbook. Vol 2: Lesson plans* (pp. 147–154). Amer Library Association.
- Bhatia, R. (2018, November). Why it is important for students to get into data literacy at an early stage. <https://analyticsindiamag.com/why-it-is-important-for-students-to-get-into-data-literacy-at-an-early-stage/>
- Biesta, G. J. (2004). Education, accountability, and the ethical demand: Can the democratic potential of accountability be regained? *Educational Theory*, 54(3), 233–250. <https://doi.org/10.1111/j.0013-2004.2004.00017.x>
- Biesta, G. J. (2012). Giving teaching back to education: Responding to the disappearance of the teacher. *Phenomenology & Practice*, 6(2), 35–49.
- Blattman, J. (2018, August). Netflix: Binging on the algorithm.
<https://uxplanet.org/netflix-binging-on-the-algorithm-a3a74a6c1f59>
- Bowler, L., Acker, A., Jeng, W., & Chi, Y. (2017). “It lives all around us”: Aspects of data literacy in teen's lives. *Proceedings of the Association for Information Science and Technology*, 54(1), 27–35.
- Bradbury, A., & Roberts-Holmes, G. (2017). *The datafication of primary and early years education: Playing with numbers*. Routledge.
- Breiter, A., & Hepp, A. (2018). The Complexity of Datafication: Putting digital traces in context. In A. Hepp, A. Breiter, & U. Hasebrink (Eds.),

- Communicative figurations: Transforming communications in times of deep mediatization* (pp. 387–405). Palgrave Macmillan.
- Buber, M. (1937). *I and thou*. T. & T. Clark.
- Center for Democracy & Technology. (2011). What does “do not track” mean? (Proposal). Washington: Center for Democracy & Technology.
<https://web.archive.org/web/20190906105058/https://www.cdt.org/files/pdfs/CDT-DNT-Report.pdf>
- Chan, T. F. (2018, May). A school in China is monitoring students with facial-recognition technology that scans the classroom every 30 seconds.
<https://www.businessinsider.com/china-school-facial-recognition-technology-2018-5?r=US&IR=T>
- Chen, M. D., Kuo, C. C., Pellegrini, C. A., & Hsu, M. J. (2016). Accuracy of wristband activity monitors during ambulation and activities. *Medicine & Science in Sports & Exercise*, 48(10), 1942–1949.
<https://doi.org/10.1249/MSS.0000000000000984>
- Cook, H. (2018, August). “It was creepy”: The parents opting out of technology in the classroom.
https://amp.theage.com.au/national/victoria/it-was-creepy-the-parents-opting-out-of-technology-in-the-classroom-20180825-p4zzqf.html?_twitter_impression=true
- Cope, B., & Kalantzis, M. (2015). Interpreting evidence-of-learning: Educational research in the era of big data. *Open Review of Educational Research*, 2(1), 218–239.
<https://doi.org/10.1080/23265507.2015.1074870>
- Couldry, N., & Yu, J. (2018). Deconstructing datafication’s brave new world. *New Media & Society*, 20(12), 4473–4491.
<https://doi.org/10.1177/1461444818775968>
- Cowie, B., & Cooper, B. (2017). Exploring the challenge of developing student teacher data literacy. *Assessment in Education: Principles, Policy & Practice*, 24(2), 147–163.
<https://doi.org/10.1080/0969594X.2016.1225668>
- Durall, E., & Gros, B. (2014, April). Learning analytics as a metacognitive tool. In *Proceedings of the 6th International Conference on Computer Supported Education* (pp. 380–384).
<https://doi.org/10.5220/0004933203800384>
- Ebner, M., & Schön, M. (2013). Why learning analytics in primary education matters. *Bulletin of the Technical Committee on Learning Technology*, 15(2), 14–17.
- Edwards, R. (2015). Software and the hidden curriculum in digital education. *Pedagogy, Culture & Society*, 23(2), 265–279.
<https://doi.org/10.1080/14681366.2014.977809>
- Erstad, O., & Voogt, J. (2018). The twenty-first century curriculum: issues and challenges. In: Voogt J., Knezek G., Christensen R., Lai KW. (Eds.), *Second handbook of information technology in primary and secondary Education* (19–36). Springer.

- Ervasti, M., Kinnula, M., & Isomursu, M. (2010). User experiences with mobile supervision of school attendance. *International Journal on Advances in Life Sciences*, 2(1–2), 29–41.
- Esmaili, D. Z., Mohamadrezai, H. & Mohamadrezai, A. 2015. The role of teacher's authority in students' learning. *Journal of Education and Practice*, 19(6), 1–15.
- Fitbit (n.d.). Fitbit.
<https://play.google.com/store/apps/details?id=com.fitbit.FitbitMobile&hl=fi>
- Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do*. Morgan Kaufmann Publishers.
- Gebre, E. H. (2018). Young adults' understanding and use of data: Insights for fostering secondary school students' data literacy. *Canadian Journal of Science, Mathematics and Technology Education*, 18(4), 330–341.
<https://doi.org/10.1007/s42330-018-0034-z>
- Goodyear, V. A., Kerner, C., & Quennerstedt, M. (2019). Young people's uses of wearable healthy lifestyle technologies: Surveillance, self-surveillance and resistance. *Sport, Education and Society*, 24(3), 212–225.
<https://doi.org/10.1080/13573322.2017.1375907>
- Giroux, H. A., & Penna, A. N. (1979). Social education in the classroom: The dynamics of the hidden curriculum. *Theory & Research in Social Education*, 7(1), 21–42.
<https://doi.org/10.1080/00933104.1979.10506048>
- Harwell, D. (2018, June). Unproven facial-recognition companies target schools, promising an end to shootings.
https://www.washingtonpost.com/business/economy/unproven-facial-recognition-companies-target-schools-promising-an-end-to-shootings/2018/06/07/1e9e6d52-68db-11e8-9e38-24e693b38637_story.html
- Ifenthaler, D., & Schumacher, C. (2016). Student perceptions of privacy principles for learning analytics. *Educational Technology Research and Development*, 64(5), 923–938. <https://doi.org/10.1007/s11423-016-9477-y>
- Jarke, J., & Breiter, A. (2019). The datafication of education. *Learning, Media and Technology*, 44(1), 1–6.
<https://doi.org/10.1080/17439884.2019.1573833>
- Kentli, F. D. (2009). Comparison of hidden curriculum theories. *European Journal of Educational Studies*, 1(2), 83–88.
- Kitchin, R. (2014). *The data revolution*. Sage.
- Koltay, T. (2015). Data literacy: In search of a name and identity. *Journal of Documentation*, 71(2), 401–415. <https://doi.org/10.1108/JD-02-2014-0026>
- Kurvinen, E., Kaila, E., Kajasilta, H., & Laakso, M. J. (2019, May). Teachers' Perceptions of Digital Learning Path in Mathematics, Languages and Programming. In *Proceedings of 242nd International Convention on*

- Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 643–648). IEEE. http://docs.mipro-proceedings.com/ce/39_ce_5294.pdf
- Lewis, S., & Holloway, J. (2019). Datafying the teaching ‘profession’: Remaking the professional teacher in the image of data. *Cambridge Journal of Education*, 49(1), 35–51.
<https://doi.org/10.1080/0305764X.2018.1441373>
- Lingard, B., Martino, W., & Rezai-Rashti, G. (2013). Testing regimes, accountabilities and education policy: Commensurate global and national developments. *Journal of Educational Policy*, 28(5), 539–556.
<https://doi.org/10.1080/02680939.2013.820042>
- Lupton, D., & Jutel, A. (2015). “It’s like having a physician in your pocket!” A critical analysis of self-diagnosis smartphone apps. *Social Science & Medicine*, 133, 128–135.
<https://doi.org/10.1016/j.socscimed.2015.04.004>
- Lupton, D., & Williamson, B. (2017). The datafied child: The dataveillance of children and implications for their rights. *New Media & Society*, 19(5), 780–794. <https://doi.org/10.1177/1461444816686328>
- Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher*, 42(1), 30–37. <https://doi.org/10.3102/0013189X12459803>
- Markham, A. N. (2020). Taking data literacy to the streets: Critical pedagogy in the public sphere. *Qualitative Inquiry*, 26(2), 227–237.
<https://doi.org/10.1177/1077800419859024>
- Mascheroni, G. (2018). Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology*.
<https://doi.org/10.1177/0011392118807534>
- Merikivi, J., Myllyniemi, S., & Salasuo, M. (2016). *Media hanskassa. Lasten ja nuorten vapaa-aikatuiskimus [The leisure time study of children and youth]*. Nuorisotutkimusseura.
- O’Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Penguin.
- Paakkari, A. (2020). *Entangled Devices: An ethnographic study of students, mobile phones and capitalism*. Doctoral dissertation, University of Helsinki, Finland.
<https://helda.helsinki.fi/bitstream/handle/10138/310928/Entangle.pdf?sequence=1&isAllowed=y>
- Paananen, M. (2017). *Imagaries of early childhood education: Societal roles of early childhood education in an era of accountability*. Doctoral dissertation, University of Helsinki, Finland.
<https://helda.helsinki.fi/bitstream/handle/10138/173807/Imaginarie.pdf?sequence=1&isAllowed=y>
- Palsa, L., & Mertala, P. (2019). Multiliteracies in local curricula: conceptual contextualizations of transversal competence in the finnish curricular

- framework. *Nordic Journal of Studies in Educational Policy*, 5(2), 114–126. <https://doi.org/10.1080/20020317.2019.1635845>
- Pangrazio, L., & Sefton-Green, J. (2019). The social utility of “data literacy”. *Learning, Media and Technology*. Advanced online publication. <https://doi.org/10.1080/17439884.2020.1707223>
- Pangrazio, L., & Selwyn, N. (2018). “It’s not like it’s life or death or whatever”: Young people’s understandings of social media data. *Social Media + Society*, 4(3), 1–9. <https://doi.org/10.1177/2056305118787808>
- Pangrazio, L., & Selwyn, N. (2019). “Personal data literacies”: A critical literacies approach to enhancing understandings of personal digital data. *New Media & Society*, 21(2), 419–437. <https://doi.org/10.1177/1461444818799523>
- Pangrazio, L., & Selwyn, N. (2020). Towards a school-based ‘critical data education’. Advanced online publication. *Pedagogy, Culture & Society*. <https://doi.org/10.1080/14681366.2020.1747527>
- Pijera-Díaz, H. J., Drachsler, H., Järvelä, S., & Kirschner, P. A. (2019). Sympathetic arousal commonalities and arousal contagion during collaborative learning: How attuned are triad members? *Computers in Human Behavior*, 92, 188–197. <https://doi.org/10.1016/j.chb.2018.11.008>
- Polar (n.d.). Ota pikaharppaus liikuntakasvatuksessa tekniikkamme avulla [improve your physical education classes with our technology]. https://www.polar.com/fi/b2b_tuotteet/liikuntakasvatus
- Portelli, J. P. (1993). Exposing the hidden curriculum. *Journal of Curriculum Studies*, 25(4), 343–358. <https://doi.org/10.1080/0022027930250404>
- Schildkamp, K., Poortman, C. L., & Handelzalts, A. (2016). Data teams for school improvement. *School effectiveness and school improvement*, 27(2), 228–254. <https://doi.org/10.1080/09243453.2015.1056192>
- Song, Y. (2014). “Bring your own device (BYOD)” for seamless science inquiry in a primary school. *Computers & Education*, 74, 50–60. <https://doi.org/10.1016/j.compedu.2014.01.005>
- Spina, C. (2017, June). Why kids need data literacy, and how you can teach it. <https://www.slj.com/?detailStory=why-kids-need-data-literacy-and-how-you-can-teach-it>
- Špiranec, S., Kos, D., & George, M. (2019). Searching for critical dimensions in data literacy. *Information Research*, 24(4). <http://informationr.net/ir/24-4/colis/colis1922.html>
- Statista (2020, February). Global unit shipments of sports watches 2017-2019, 2022. <https://www.statista.com/statistics/385753/sports-watches-worldwide-shipments/>
- Raviv, A., Bar-Tal, D., Raviv, A., Biran, B., & Sela, Z. (2003). Teachers’ epistemic authority: Perceptions of students and teachers. *Social Psychology of Education*, 6(1), 17–42. <https://doi.org/10.1023/A:1021724727505>

- Rytönen, A-P. (2019, November). Kännyköihin ladataan kouluissa sovelluksia, jotka äänittävät lasten puhetta—Isä vaatii: “Tällaisia tietoja ei saa luovuttaa ulkopuolisille” [Schools download applications that record children’s speech—Father insists: “Such information must not be passed on to outsiders”]. <https://yle.fi/uutiset/3-11046004>
- Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. *Big Data & Society*, 6(1).
<https://doi.org/10.1177/2053951718820549>
- Selwyn, N. (2018). Data points: Exploring data-driven reforms of education. *British Journal of Sociology of Education*, 39(5), 733–741.
<https://doi.org/10.1080/01425692.2018.1469255>
- Selwyn, N. (2019a). *Should robots replace teachers? AI and the future of education*. Polity.
- Selwyn, N. (2019b). What’s the problem with learning analytics? *Journal of Learning Analytics*, 6(3), 11–19.
<http://dx.doi.org/10.18608/jla.2019.63.3>
- Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78, 397–407.
<https://doi.org/10.1016/j.chb.2017.06.030>
- Siljander, P. (2002). *Systemaattinen johdatus kasvatustieteeseen [Systematic introduction to educational sciences]*. Vastapaino.
- Slotte Dufva, T. (forthcoming). *Post-digital art education—boundary-shifting and digi-grasping the perceived world*. Palgrave.
- Stevenson, H. (2017). The “datafication” of teaching: Can teachers speak back to the numbers? *Peabody Journal of Education*, 92(4), 537–557.
<https://doi.org/10.1080/0161956X.2017.1349492>
- Touhula (n.d.). Polar Electro: Supporting active life.
<https://touhula.fi/en/partner/polar-electro/>
- Valtonen, T., Tedre, M., Mäkitalo, K., & Vartiainen, H. (2019). Media literacy education in the age of machine learning. *Journal of Media Literacy Education*, 11(2), 20–36. <https://doi.org/10.23860/JMLE-2019-11-2-2>
- Van Dijck, J (2014) Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology. *Surveillance & Society*, 12(2), 197–208. <https://doi.org/10.24908/ss.v12i2.4776>
- Wang, F., Tong, Y., & Danovitch, J. (2019). Who do I believe? Children’s epistemic trust in Internet, teacher, and peer informants. *Cognitive Development*, 50, 248–260.
<https://doi.org/10.1016/j.cogdev.2019.05.006>
- Weiser, M. (1991). The computer for the 21st century. *Scientific American*, 265(3), 94–105.
- Williamson, B. (2017). *Big data in education: The digital future of learning, policy and practice*. Sage.
- Williamson, B. (2017b). The digitised future of physical education: Activity trackers, biosensors and algorithmic biopedagogies. In M. Thornburn

- (Ed.), *Transformative learning and teaching in physical education* (pp. 176–192). Routledge.
- Wilson, P. (1983). *Second-hand knowledge: An inquiry into cognitive authority*. Greenwood Press.
- Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics*, 12(3), 9–26.
- Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, 129, 41–54. <https://doi.org/10.1016/j.ijhcs.2019.03.006>