

# Datafication in and of Education – a literature review

DELIVERABLE D4.6

Ola Erstad, Øystein Gilje, Greta Björk Gudmundsdottir, Kristiina Kumpulainen, Olga Viberg, Ben Williamson, Jo Tondeur and Sarah Howard

30 June 2025

# Contents

Acknowledgements .....	4
1. Introduction .....	5
2. Key terms.....	8
2.1. Inclusion and equity .....	8
2.2. Data, datafication, platformisation and data literacy.....	9
2.3. Big data and learning analytics.....	11
2.4. Artificial intelligence and machine learning .....	11
2.5. Other issues – implications.....	12
3. Methods and scope .....	14
3.1. Search for empirical research .....	14
3.2. Other research publications .....	16
4. Findings from the literature.....	18
4.1. Data Regulation, Rights and Privacy.....	18
4.2. Data in Use for Teaching and Learning.....	22
4.2.1. Learning analytics .....	22
4.2.2. The Importance of Data Literacy and Data Agency .....	24
4.2.3. Teacher’s professional development.....	25
4.3. Data Governance .....	27
4.3.1. Education governance through data.....	28
4.3.2. Data-driven digital governance.....	29
4.3.3. Platform governance and educational autonomy.....	30
4.3.4. Governing stacks and creeping commercialization.....	31
4.3.5. Automated governance and pre-emptive judgement.....	32
4.3.6. Data ownership .....	33

4.3.7. Governance of data .....	34
5. Towards an Analytical Framework.....	37
6. Defining key themes and questions .....	41
7. Education Data Futures .....	43
8. References .....	45

# Acknowledgements

Work on the literature review has been an ongoing process throughout the three-year period of the Agile-EDU project. A first draft was presented by the authors in June 2023. At that time, it was important to define an analytical model based on a review of relevant literature on datafication and education, presented at the end of the literature review, because it has been important in defining key perspectives in the selection of case studies, learning stories and different topics within the scope of the project.

Since the first draft was presented the different partners in the project have provided input to the text. We also organized an online seminar with invited scholars and stakeholders where the first draft and the analytical model were discussed. The feedback provided valuable input.

Thank you to the secretariat at European Schoolnet for providing critical and constructive comments that helped developing the text further. We also want to thank Rebekka Baunbæk Wagstaffe for her work in doing the initial review of literature in different databases and writing the methodology.

The work has been organised with several online meetings, especially in the first phase of writing this report, and with several revisions and input from authors until June 2025.

We hope this work can be helpful for both policy makers, practitioners and researcher within this emerging field of great importance for our societies and education systems.

# 1. Introduction

As European citizens we live in technology-rich societies, where handling of digital data has become a key economic driving force in what is often termed as the 'attention economy' (van Krieken, 2018) and 'surveillance capitalism' (Zuboff, 2019). The foundational element is digital data, which has had a major impact on our conceptions of socio-cultural developments, health management systems and our education systems. Policy makers, regulators, researchers and practitioners express both visionary prospects and fundamental concerns about these developments, primarily due to the rise of global technology companies and the ways they have entered education systems on a scale never seen before.

Digital data is embedded in every aspect of life, from health tracking apps, social media, learning management systems, to the Internet of Things, traffic surveillance, government services, dating sites and many more (Pangrazio & Sefton-Green, 2022). Most of our online activities are registered and stored as digital data, as we consciously or unconsciously accept and create our own digital footprints. Increasingly, the public and policymakers are becoming aware that all these data are collected and may be used by others somewhere else in the world for different purposes (Hakimi et al., 2021).

Furthermore, the Covid-19 pandemic triggered a whole new setting for digital education and the provision and use of digital data for improved decision-making. Education is of special interest concerning the handling of data and datafication of specific sectors in society due to children and young people's status as vulnerable participants in society (Nottingham et al., 2022). The ways data-driven technologies have entered education during the last twenty years have created new challenges and possibilities. Moreover, the data collected from children through their participation in school are exponentially increasing in variety and volume (Livingstone & Pothong, 2022, p. 17). The collection of data implies documenting students' progress, often in numerical forms, as a mechanism of control (Williamson, 2017).

This literature review is part of the Erasmus+ Forward Look project 'Agile-EDU', led by European Schoolnet. The project lasts from January 2023 until December 2025. A starting point for the project is to identify the key studies that can define an analytical framework to be used for the project. However, this review is important in itself in that it brings together literature and research in this field in order to highlight some key issues and questions, with a distinct European focus on an emerging area of interest for many stakeholders – that of data, datafication and education. The people who have collaboratively written this report are either partners in the project from the Faculty of Education, University of Oslo (Ola Erstad, Øystein Gilje, Greta Björk Gudmundsdottir, Rebekka Baunbæk Wagstaffe) or members of the Scientific Advisory Board for the project

(Kristiina Kumpulainen, Olga Viberg, Ben Williamson, Jo Tondeur, with additional contributions from Sarah Howard).

The aim of the Agile-EDU project is to clarify and explore some of the key terms in the field of datafication and education. The focus is on meaningful and responsible use of data, all framed within conceptions of inclusiveness, equality and equity in contemporary school ecosystems in the age of datafication. The question guiding the work in this literature review is:

*What are the conditions and criteria for responsible, inclusive and meaningful use of data to support teaching and learning?*

Responsible, inclusive and meaningful use of data to support teaching and learning refers to elements that are external to classroom activities, such as regulations of ethical norms and privacy. Moreover, key questions in the literature are also related to governance. These two topics have an impact on teaching and learning, but are analytically distinct and will thus be presented in separate sections in this report. The major concern, and a major theme in the literature, is the use of data as part of teaching and learning in the classroom (Pangrazio & Sefton-Green, 2022). Consequently, the review addresses three themes (1) Data regulation, rights and privacy (2) Data in use for Teaching and Learning and (3) Data governance. Across the different sections we highlight what is at stake concerning the datafication of education and contemporary developments of the whole ecosystem involving teaching and learning in schools. Ecosystem in this context implies an examination of the multivalent and mutually constitutive relationships that link surroundings to that which they surround (Benson, 2020; Stornaiuolo, Higgs, Nichols, Leblanc & de Roock, 2023: 313). A key question, as stated by Livingstone and Pothong (2022, p. 17), to be aware of, is; “Whose interests are served by the intensifying ‘datafication’ of education and childhood?”

The GDPR (General Data Protection Regulation) and the AI Act has been implemented in Europe to regulate companies’ access to digital data. Europe is of special interest in this context, both because of the implementation of these regulatory measures and an emerging awareness within European education systems of the potentials and risks of digital data ecosystems. The AI Act (August 2024) is a European regulation on artificial intelligence (AI) – the first comprehensive regulation on AI by a major regulator anywhere. The Act assigns applications of AI to three risk categories. First, applications and systems that create an unacceptable risk, such as government-run social scoring of the type used in China, are banned. Second, high-risk applications, such as a CV-scanning tool that ranks job applicants, are subject to specific legal requirements. Lastly, applications not explicitly banned or listed as high-risk are largely left unregulated.

Both the project and this review address a specific European agenda on data and the datafication in and of education. In the review we also draw on literature and research from other parts of the world, especially Australia as they have several leading research communities in this field of research and are comparable with the European context.

The major challenge of writing a literature review in this field is of course the speed of change of technologies in later years and the expanding number of publications that have become available

in a rather short period of time. It seems something new is being published every day. Hence, we have had to limit our ambitions of what to include, and this report is not providing a complete overview of literature in the field.

This review is first and foremost intended to a) clarify concepts and provide operational definitions; b) identify existing evidence from research; and c) define the conceptual framework for the Agile-EDU project. The review is intended for policy makers, school principals, teacher educators, teachers and other stakeholders interested in the topic, and is structured as follows: after this introduction we present key terms relevant for this report, and then the method and scope of the review. In section four we present the main findings from the literature leading to an analytical framework as well as key themes and questions. Finally, some reflections concerning educational data futures are presented in the end of the report.

## 2. Key terms

The terminology that has evolved on the implementation of data-driven technologies in primary and secondary schools is complex and difficult to clearly define. There is a wide variety of terms being used, all of them coined in specific discourses across a variety of domains and theories. Hence, one single term is often interpreted and used in different ways by stakeholders, policymakers, practitioners and researchers. Below, we present a short explanation of how we use and interpret the main terms of relevance in this review.

### | 2.1. Inclusion and equity

A guiding perspective both for the project and this review is the ways in which inclusion, equality and equity are defined. A key goal of educational institutions around the world is to provide inclusive and equitable education, achieving of which requires contextualised approaches to accommodate diverse global values and promote learning opportunities that best meet learners' needs (Viberg et al., 2024), and provide good education in the age of measurement (Biesta, 2009). These key terms in our understanding of datafication and education, concern the role of digital data and how they are gathered, interpreted and used throughout primary and secondary schools in Europe.

Traditionally, issues of inclusion and equity have been defined in accordance with access and equal possibilities with schooling, teaching and learning (See, e.g., Tondeur et al., 2011). And not all municipalities, schools or families can afford the same digital resources and infrastructure as others, creating digital divides. Statistics in most countries document how students' socioeconomic background defines their success at school, often raising issues of class difference (Livingstone et al., 2021). For example, artificial intelligence (AI) technologies used in education can be beneficial for students, teachers and school principals, but to ensure that such AI-based systems for improved decision-making in education advance equity and inclusion for all students, we have to evaluate these systems and tools comprehensively, since AI technologies also pose risks (Yan et al., 2024) such as they have the potential to perpetuate and exacerbate inequalities due to inherent biases in training datasets, their potential to influence real-world decision making (Viberg et al., 2024) such as automated grading.

An important question is how this manifests itself when using digital data as part of teaching and learning in schools. Will the use of data provide other means of fair treatment, full participation and inclusion of all, or does it just create new means of control and inequality among students?

In her book 'The Digital Disconnect. The social causes and consequences of digital inequalities' (2021), Ellen Helsper discusses three aspects of conceptions of inequality; access, skills and use,

and outcomes, and how these terms have moved towards increased complexity. She states that; “This complexity of all aspects of digital inequality demands care to be taken when they are operationalized for research or tackled through interventions.” (2021: 29).

In recent years terms like adaptive or personalised learning have been used to give the impression that all students will be included and taught adapted to their needs (Faber et al., 2017; Williamson, 2017; Kippers et al., 2018). One might ask if this is correct or just a new way to approach familiar issues of inclusion and equity? On a fundamental level it is important to understand how the use of data in our societies are expressions of power and surveillance, and what is referred to as dataveillance (Lupton & Williamson, 2017). We will explore some of these issues based on the research literature below.

More recent contributions, and also of more immediate interest to the issues raised in this report, is Kizilcec and Lee’s (2022) work on ‘algorithmic fairness in education’ and Baker and Hawn’s (2021) work on ‘algorithmic bias in education’. They refer to concerns about the fairness of the predictions and uses of these algorithmic systems (measurement, model learning, and action) to identify sources of bias and discrimination in the process of developing and deploying such systems. Primarily related to artificial intelligence in education, but also other sectors of society, there are concerns about the substantial impacts of algorithmic bias, which groups are known to be impacted and which stages and agents in developing and deploying educational algorithms are implicated (Kordzadeh & Ghasemaghaei, 2021). In a recent special issue on advancing equity and inclusion in educational practices with AI-powered educational decision support systems, some scholars highlight that when “focusing on equity, it is also imperative to define what we mean by this term. Equity goes beyond fairness in algorithmic models; it involves ensuring that the decisions supported by AI technologies promote inclusive and just outcomes for all learners” (Viberg et al., 2024, p. 1978).

## 2.2. Data, datafication, platformisation and data literacy

Two kinds of **data** are routinely collected in education: personal data and process data. While “personal data refers to any information that can be directly attributed to an individual” (Pangrazio & Sefton-Green, 2022: 2), process data refers to any information collected from a user’s interaction with software, such as clicks and keystrokes on the interface, and time spent on a particular activity, documents or materials accessed, videos viewed, and so on. Both types of data are gathered in diverse settings and ways through our everyday activities, also as part of digitally processed data about student learning.

Sometimes we are aware of data creation, like making a profile, while other times we are not. However, as Pangrazio and Sefton-Green emphasise, part of the problem is that the term digital data “is a rather nebulous concept that refers to so many seemingly different pieces of information.” (ibid.). In this review, we understand student’s education data as all data gathered

from and about students and their activities during their participation in formal and informal learning.

**Datafication** is the process through which actions, interactions and behaviours are translated into digital data that can be collected, sorted, analysed or commodified by governments and private companies (Pangrazio & Sefton-Green, 2022). In education, datafication occurs through the government collection of educational records and the production of large regional or national databases. In addition, the collection of information by education technology providers is also considered as datafication in and of education in this review. The data can be subjected to processing, analysis, interpretation, visualisation and circulation using a wide variety of technologies and methods, particularly data analytics and other techniques derived from data science and statistics.

A related concept used in the literature on datafication is 'data assemblages' understood as "the technological, political, social and economic apparatuses and elements that constitute and frame the generation, circulation and deployment of data" (Iliadis & Russo, 2016: p. 1). This term refers to the whole ecosystem of data infrastructures as well as people, processes and practices of use, for example within social media platforms such as Instagram or an internet connected baby monitor that create assemblage of data when used.

Datafication in and of education is driven by different purposes and by diverse agents. For government agencies, large quantities of digital data can be used for monitoring whole educational systems (for example, national assessment data) and for auditing the performance of individual schools. Consequently, datafication is seen as a route to improvements in learning, learning outcome, teaching and school management and administration. For the providers of the systems to store the data, like the EdTech companies, data is also valued as a source to be used in order to create new products and services that can be marketed to schools and therefore 'monetized' (Williamson, 2017). The datafication of education has been further advanced by the education technology industry with the appearance of digital platforms.

**Platformisation** refers to the widespread introduction of digital platforms into many aspects of everyday life, spanning social interaction and communication (social media platforms), online commerce, travel and transport (smartphone taxi booking services), and public services ('Government as a Platform') (Poell et al., 2019). Platforms are understood as key drivers of the global economy, often termed 'platform capitalism' (Srnicsek, 2017). Central to the platform model is the capacity for the service to run online, be modifiable and updateable, generate income from ongoing subscriptions or premium service payments, and continuously capture data about how they are used. This may lead to a monopolisation as many of the providers of platforms "lock" the data into their specific system, with limited solutions for interoperability (Kerssens & van Dijck, 2022).

Digital education platforms have become increasingly common in schools. The software systems routinely used by schools, such as learning management systems, digital textbooks, digital learning services and other forms of educational technology, often now take the form of a

platform, and can therefore generate very large quantities of digital data when students and teachers interact with them (Perrotta et al., 2020). Platformisation of education is therefore tightly coupled to increasing datafication of education, particularly as platforms can generate new sources of information for analysis, and the data can be used by the proprietor to develop new services and upgrades.

**Data literacy** is an important term, and we will use it primarily for understanding how teachers need to develop new digital competences in their professional learning. Although, we are well aware that also students need to be supported in their data literacy and use (Wasson & Kirschner, 2020). Data literacy for teachers is the ability to transform data into actionable instructional knowledge and practices.

### | 2.3. Big data and learning analytics

Authorities have collected educational information and records for over a century, but datafication and platformisation increasingly involve the collection and processing of very large quantities of 'big data' that consist of both personal student data and process data about students' interactions with software systems, as well as teachers understanding and use of these data in their work (Williamson, 2017). These forms of data, often combined, are said to make it possible to generate novel insights about students' engagement, learning and academic performance, even to generate predictions of their likely future progress and outcomes. Such big data can be used as 'actionable intelligence' and the basis for various forms of pedagogic data-driven decision-making, and potentially even automated 'prescriptive analytics' interventions or 'adaptive' forms of 'personalised learning' (Selwyn et al., 2023). The academic research and the developments in the fields of Learning Analytics and Educational Data Mining have provided the intellectual and technical basis for the use of big data in education. Learning analytics (LA) is about measuring, collecting, analysing and reporting data to improve students' learning experiences and to support learning and the environments in which it occurs (Siemens & Long, 2011). Data can be used to provide detailed feedback on students' performance and provide new insights into the learning process through learning analytics (Gasevic et al., 2015). Thus, we understand the use of big data and LA in this review as an endeavour that is an established research field in other education settings like higher education, but rather new in K-12 research academic settings. As a field of research, it is a specific practice driven by the digital platforms built by the EdTech industry, as well as some open source platforms like Moodle, and used in primary and secondary education.

### | 2.4. Artificial intelligence and machine learning

The rising use of big data and analytics in education has paved the way for the emergence of artificial intelligence (AI) in education (Bill et al., 2021; Carvalho et al., 2022). AI is a highly contested term, as it is often used to refer to a wide variety of technologies, from those that deploy relatively simple algorithms in intelligent tutoring systems (ITS) to far more sophisticated data-

driven systems that can interpret and process text and images (or 'generative AI'). This leads to a whole range of ethical issues, as well as copyright issues with AI-generated text and images being included in schoolwork as well as published on the web. In the school year 2022/2023, recent development of AI has also been the source of both immense hype and anxiety, in particular in relation to natural language processing in use by generative AI tools such as ChatGPT and BARD (Malinka et al. 2023).

Within the educational context, AI actually has a long history stretching back to the 1960s, in the shape of the academic research and development field of 'AIED'. With the availability of educational big data and advanced algorithmic techniques known as 'machine learning', AI is now routinely seen as a transformative technology in education, although there remains limited evidence of its effects (either intended or not) (Holmes et al, 2022). Machine learning refers to a particular kind of algorithm that can be 'trained' on a sample of data to then find patterns in other data sources, or 'learns' from the data it processes to then generate an output, rather than a classical data-processing algorithm that has to be programmed to look for specific features in data. Put in an educational context, machine learning underpins AI applications that can process very large quantities of student data, identify patterns, make predictions, and potentially even prescribe interventions (Perrotta & Selwyn, 2020). New services like ChatGPT and BARD have used machine learning to train large language models (LLMs), that easily generate texts based upon students and teachers prompts.

Other relevant aspects like computational thinking and data literacy (see above), as well as privacy- and AI literacy, have become important in conceiving the future of education. In 2024, UNESCO (2024) released two comprehensive frameworks that guide teachers (<https://www.unesco.org/en/articles/ai-competency-framework-teachers>) and students (<https://www.unesco.org/en/articles/ai-competency-framework-students>) on AI use and misuse in education, defining the knowledge, skills, and values they must master in the age of AI. For teachers, specifically, the three progression levels (i.e., acquire, deepen and create) are specified in regard to the acquisition and development of 15 competences across five dimensions: Human-centred mindset, Ethics of AI, AI foundations and applications, AI pedagogy and AI for professional learning.

## | 2.5. Other issues – implications

Our focus in this literature review is oriented towards what research in the field of datafication of and in education can tell us about responsible, inclusive and meaningful use of data to support teaching and learning. It is beyond our scope to cover all possible issues of relevance in this diverse and emerging field. However, based on our review, we raise some themes as implications and output towards the end, and which will be embedded in different sections of this report. These themes include questions like:

- What competencies are important for teachers?

- How will teacher training and teacher's professional practice change due to datafication?
- What type of support is available for teachers?
- How does platformisation influence the context for learning?

Finally, ethical considerations, privacy and trust are inevitably of concern when using huge amounts of data in teaching and learning in schools.

Some of these questions are covered in the research literature, but very little is based on empirical data and research in primary and secondary schools. However, the implications of input from the research literature might enlighten how these important questions and issues can be worked on. We return to these challenges towards the end of this review.

# 3. Methods and scope

This report is neither a systematic nor a scoping review. The time constraints of writing the report have made us focus on a selected number of publications to identify central issues of relevance for the Agile-EDU project. In order to organise our work, we have used different approaches to select and analyse what we have considered as relevant research literature. This implies that there might be literature that we do not cover or refer to in this report that still could be important and relevant.

Our aim with this report is to specify some key areas of importance and issues that need more attention both by policy makers, practice and research, defined as an analytical framework. As a literature review it has been important to identify some key literature and areas that need attention in the field and not present this as a field of research in itself. Still, we have tried to build our report on systematic literature searches and our own expertise and networks within this field. Some criteria for how this has been done are presented below.

## 3.1. Search for empirical research

The first part of our review is based on a systematic search for published empirical articles in international journals. The search string developed aimed at finding studies that could inform us about empirical articles addressing the three areas guiding our work.

- Data, regulation, rights and privacy (see Section 4.1)
- Data in use for teaching and learning (see Section 4.2)
- Data governance (see Section 4.3)

Our first step was to define search strings that could help us in identifying relevant publications and empirical studies. Each of the three areas had their own search string, but the same search strategy was applied on each area. Table 1 shows the search strings used, and how many articles were found for each area in the first round.

As mentioned above, we are not conducting a systematic review that would imply selecting and coding empirical studies in specific ways. Rather, we have used a systematic search to identify key literature and then read across a number of selected whole text articles to define some common themes of importance.

AREA 1: Regulation and rights	AREA 2: Teaching and learning	AREA 3: Governance
((("data" OR "datafication" OR "data education" OR "artificial intelligence" OR "platforms" OR "learning analytics") AND ("regulation" OR "privacy" OR "safety" OR "ethics"))	((("data" OR "datafication" OR "data education" OR "artificial intelligence" OR "platforms" OR "learning analytics") AND ("teaching" OR "learning" OR "student" OR "teacher" OR "educational assessment" OR "digital literacy"))	((("data" OR "datafication" OR "data education" OR "artificial intelligence" OR "platforms" OR "learning") AND ("governance" OR "ownership" OR "ecosystem"))
220 articles	634 articles	36 articles

**Table 1: Search strings and number of results from selected databases**

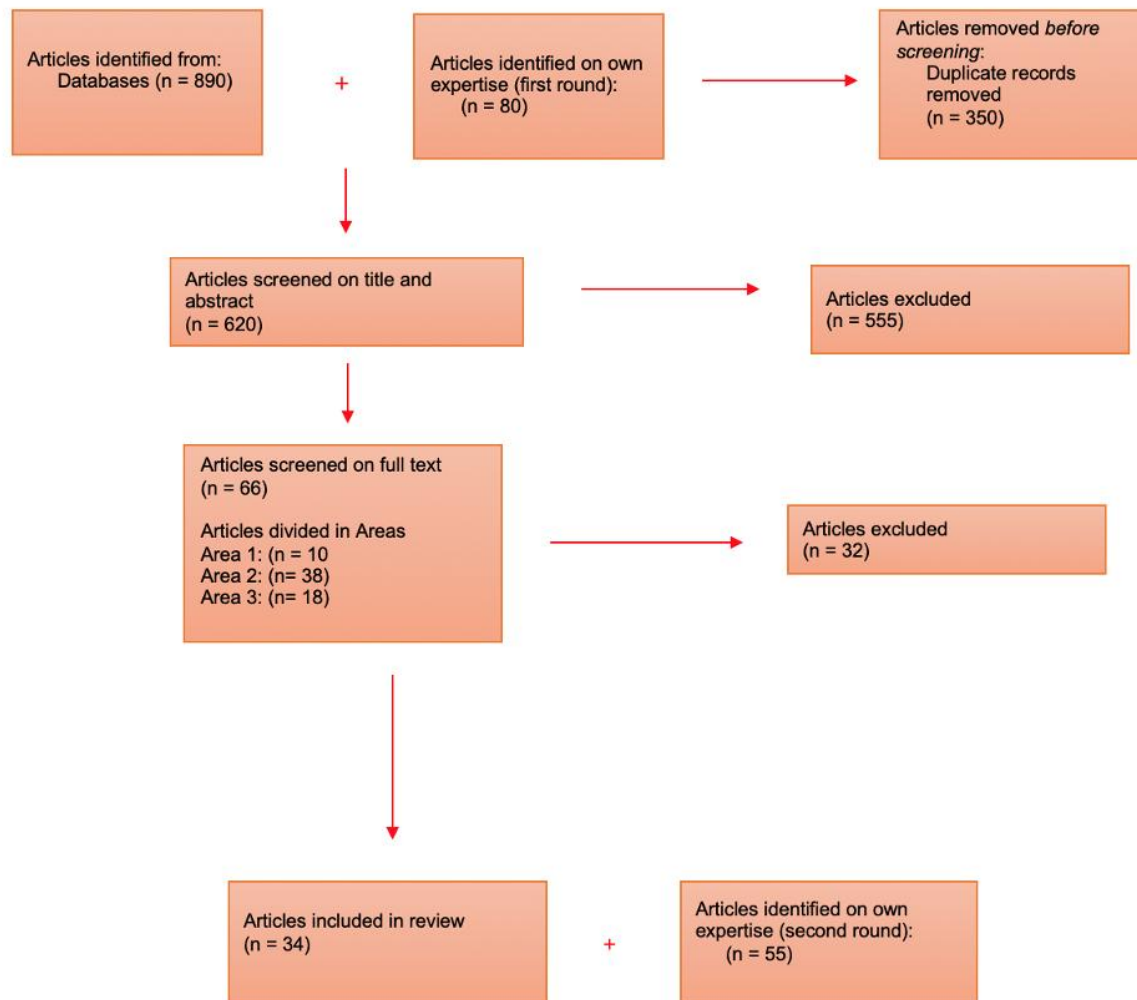
The searches were done in four selected databases: ERIC (EBSCO), ERIC (OVID), PsycINFO (OVID) and Web of Science. The limitation for each search was that the terms in the search string had to appear in the title.

To identify relevant studies some inclusion and exclusion criteria were applied. Studies were included if they involve K-12 education and teacher education in a European context, describe empirical research, were peer-reviewed, are written in English, and published between 2013 and 2023. We restricted the search to 10 years, building on our understanding of the field and previous reviews. Studies were excluded if the study interest was on countries outside of Europe, if the studies had a non-educational focus or focus on higher education, articles with a theoretical approach with no presentation of empirical evidence, and studies published 2012 or earlier.

Both the results from the search (890 articles) and also 80 articles that we had identified and entered into a Google Drive were transferred to the reference-manage system Zotero and divided into three folders representing each area. Zotero is a software for managing references and it was used both to discover duplicates and later to transfer the articles into Covidence, our software used for the systematic work of inclusion and exclusion of articles. The result was that 620 articles were transferred to Covidence (350 duplicates removed). Our experience from this first step was that the terms used were quite broad, especially for area 2 on teaching and learning. This meant that we found a substantial number of studies.

The second step was to define the relevance of each article. First, we did a screening of the titles and abstracts. Two members of the team had to choose either to include or exclude for each article independent of each other. Where there was a conflict a third person had to make a final decision in Covidence. In the software Covidence each article was categorised by tagging them with the associated area. After screening all the abstracts and titles, we ended up with the ones most relevant for our review.

The third step was to read the full-text version of the included articles. These were 10 articles in Area one (4.1), 38 in Area two (see 4.2) and 18 in Area three (4.3). The team divided the reading of articles among themselves. We did not code each article but read across all to identify themes and topics as well as results of relevance for our work. This is summarised in section 4 below.



### 3.2. Other research publications

In addition to the search and review done of published empirical research in this area, we have based our literature review on search for other publications of key importance. During the last five years several books and reports of key importance have been published in Europe and these are also included in our review, such as the 'Artificial intelligence and Education' report (Holmes et al., 2022) by the Council of Europe, 'The Nordic AI and data ecosystem 2022' report by Nordic Innovation, and work done by the Digital Futures Commission in the UK. In addition, we have identified reports by international organisations like UNESCO's report 'Minding the data: protecting learners' privacy and security' (2022). However, there is a fast-growing number of

reports in this area, many related to EU policy initiatives, which is beyond the scope of this report. We have included a few we believe is of relevance for the perspectives included in this report. Several international books have also been of relevance for our work, such as 'Learning to live with datafication. Educational case studies and initiatives from across the world' (Pangrazio & Sefton-Green, 2022). In addition, we identified and considered interesting perspectives in books like 'AI in Learning: Designing the Future' (Niemi et al, 2022) and 'The datafication of Education' (Jarke & Breiter, 2020).

As mentioned before, we concentrate our focus on a European context. That means that we primarily draw on literature within a European context. However, we also draw on some literature from outside of Europe, especially from some research communities in Australia, partly because they are of direct relevance to our work, and partly because the context within which they are written is similar to a European agenda in this field, with scholars such as Neil Selwyn, Julian Sefton-Green, Luci Pangrazio, Dragan Gašević and others.

Having done the initial search for literature, our Scientific Advisory Board participated in scanning the literature. Our Scientific Advisory Board is put together by experts in the field and each of them has a good overview of the literature included in this report, and have also themselves contributed with key publications on the topic of this literature review.

# 4. Findings from the literature

In this section we present findings from our review of literature. The three identified areas refer to key issues raised in the proposal for the Agile-EDU project concerning responsible, inclusive and meaningful use of data to support teaching and learning. One area (4.1) is about the rights and regulations set in place on European and national levels in order to manage developments in the datafication in and of education. A second area (4.2), with publications about the educational practices of using data in school settings for teaching and student learning. The third area (4.3) addresses issues of governance of data in education which is important in understanding mechanisms of how data is managed by different stakeholders and interests within the ecosystem of education.

Our main focus is on K-12 education within a European context. Our reflections and discussions might be of relevance to outside Europe, but the aim of our work is to address developments and challenges within this region in particular.

## 4.1. Data Regulation, Rights and Privacy

Digital technology – including learning analytics- and AI-based educational tools – has been increasingly adopted by schools in the setting of K-12 education (Crompton et al., 2022; Holmes, 2022; Vezzoli et al., 2020; Yu & Couldry, 2022). The use of these technologies offers increased access to interaction- and personal student digital data that could be used to improve conditions for their learning (Milkaite et al., 2021), but also creates new risks for students' privacy (Lieberman, 2020), including concerns about how schools monitor students' technology use, what data companies get when students use their devices, and how schools use the collected data (Kumar et al., 2019). New concerns about student privacy are especially emerging in relation to the use of LA and AI-based educational technologies (e.g., Bulger et al., 2017; Kumar et al., 2019; Livingstone, 2020; Stoilova et al., 2021) that aggregate and generate various types of student data, including student sensitive data. Of the 10 empirical studies we reviewed on the topic for this section, a majority of them were concerned with. In this section, we will also draw upon a large number of studies that were identified based upon our expertise.

All this suggests that we now have access to, not only more student data, but also to more nuanced data from many sources, often in real-time, allowing schools and companies (e.g., educational technology providers) detailed student profiles, creating a so-called 'datafied child'

(Lupton & Williamson, 2017; Williamson et al., 2020; Siibak & Mascheroni, 2021). A 'datafied child' is understood in terms of how intersecting monitoring platforms and devices use children's data to create digital profiles that may accompany children long after the period of collection. Children's records, including demographic, behavioural and relational data, are often stored, combined with other information sources and accompany them over the years. Such data is increasingly shared and sold to commercial interests as part of service level agreements and practices that breach students' right to privacy with possible long-term impacts (Regan & Jesse, 2019; Hooper, Livingstone & Pothong, 2022). All in all, there are different parties that have interest in using student data for various purposes. However, the question is who will protect the 'datafied' child's right to privacy in the context of K-12 education? To protect children's privacy in schools is a moral and legal obligation for school leadership, teachers and parents. Yet, there is still limited research about responsible use of students' data in schools (Kumar et al., 2019; Mikaite et al., 2021).

Scholars stress that "the status of the child online is shifting from one of invisibility to one of hypervisibility in an increasingly datafied world, and the child's right to privacy has rapidly become the most contested of all the rights" (Livingstone, 2020, p.1). Therefore, in the setting of the many challenges encountered by K-12 educational institutions, and evidence informed decision-making, the promise and concerns regarding data analytics, student agency and student privacy are pertinent. The increased datafication of K-12 education is understood in the broader context where offline and online behaviours are often converted "into online quantified data, thus allowing for real-time tracking and predictive analysis" (van Dijck, 2014, p. 198), among others.

Collecting data from students in schools through close monitoring has played a continuous role in promoting educational progress (Jenks, 2005). However, in some countries, including European countries, specific surveillance processes within schooling have attracted criticism, including school and university league tables and globally standardised tests, including PISA (the Programme for International Student Assessment), and the monitoring of children's movements and activities through CCTV cameras and radio-frequency identification chips placed on school uniforms (Yu & Couldry, 2022). Different forms of dataveillance (Lupton & Williamson, 2017; van Dijck, 2014; Yu & Couldry, 2022) entail the collection of data, increasingly real-time, from children which are stored in online corporate platforms, calculated so as to predict and manipulate future behaviour, and therefore monetised (Hintz et al., 2017; Mascheroni, 2018; van Dijck, 2014, cited in Mascheroni, 2018, p.3). Scholars argue that the introduction of surveillance techniques as tools for educational development via digital platforms has hitherto generated "little public debate", and that the emerging, data-driven discourse about education "ignores the potential costs of datafication for children, teachers and wider society [...], [in which the] agency of teacher and student in the classroom becomes displaced" (Yu & Couldry, 2022, p. 128).

Further, there is so far little evidence of specific consideration to safeguard children's rights in relation to dataveillance and propose paying attention to this (Lupton & Williamson, 2017). One of the related attempts refers to a study by Stoilova et al. (2020), in which children's (age 11-16; N = 169) understanding of privacy in three contexts: interpersonal, institutional and commercial has

been examined, in the UK school context. The results showed that children largely conceptualise privacy in relation to interpersonal settings, conceiving of personal information as something they have agency and control over as regards deciding when and with whom to share it, even though they do not always practise such control. Further, the authors argue that the complexity of the digital environment challenges teachers' capacity to address children's knowledge gaps, "businesses, educators, parents and the state must exercise a shared responsibility to create a legible, transparent and privacy-respecting digital environment in which children can exercise genuine choice and agency" (p.197). Later, in their systematic mapping of children's understanding of personal data and privacy online, Stoilova and colleagues found that most empirical studies concern children's behaviours and practices, but much less common are studies of children's media literacy, in particular linked to their capacity to consent to data protection practices (Stoilova et al., 2021).

Scholars also highlight that little research has examined how educators consider privacy and security in relation to classroom technology use (Kumar et al., 2019). Recently, Nottingham et al. (2022), in their work focusing on the exploration of the legal and ethical implication of the use of digital technologies in schools, recommend that changes in digital schooling practices are required to enable children with "realistically possible ways of enforcing their data protection rights as well as a clarified and uniformed approach to support schools" (p.1). There is also, to our knowledge, a paucity of empirical research that investigates the school leadership- and parents' perspective in this regard. All these perspectives are important to be able to define and enact a more responsible and responsible use of student data in school education settings.

Responsible use of children's data implies protection of their data privacy rights, which has received special attention in several legislative documents, such as the General Data Protection Regulation (GDPR) in the European context, the Data Governance Act (europa.eu), and also specific toolkits, such as General Principles of Children's Online Privacy and Freedom of Expression (UNESCO, 2018). Yet, these documents do not accurately reflect different stakeholders' perspectives and seldom provide practical guidelines about how they should be implemented in practice - in particular educational contexts - in which different cultural values may also play a role. For example, in the context of the Nordic countries, such values as trust, openness and transparency are ingrained in the culture at different levels, including the individuals' attitudes towards - and their acceptance of - the use of emerging digital technologies (Robinson, 2020), including AI that is nowadays increasingly used in the design of adaptive and personalised learning solutions that have been extensively adopted and used in education in different forms (for overview, see Crompton et al., 2022). Yet, most of the AI-based educational tools have not been developed with the protection of students' privacy in mind (Rauf, 2020).

There are several incitements or provocations to be considered in terms of protecting children's privacy (Livingstone, 2020). They relate to the following questions to be carefully considered: 1) Should we enable children's full participation in digital spaces, or should we minimize their risks by providing children-only, or even off-line, spaces?, 2) Should we pay from the public funds for the

provision of digital and non-digital spaces for children, or should we allow the commercialisation of these spaces?, 3) Is it an option to hold parents accountable? In this regard, Sonia Livingstone points to evidence that parents are frequently not up to the task, and that the parents of the most vulnerable are equally not prepared to take accountability. In their efforts to protect children in online spaces, they may interfere with children's rights of participation in digital spaces and, 4) To what extent can we hold digital companies responsible for the digital wellbeing of children? And even if we do, are they trustworthy? These are important questions to reflect upon in the design of relevant research studies in various cultural and socio-economical contexts, as well as in the process of examining privacy aspects in the evaluation of the digital technologies used in K-12 education settings.

In summary, there are several problems to be addressed by practitioners and researchers, and this requires the involvement of various stakeholders responsible for the protection of students' rights and privacy in schools.

The first problem relates to the fact that (educational) technology is often seen as a great equaliser for different student populations and geopolitical contexts (Prinsloo, 2018). Governments, educational departments and educators worldwide are encouraged and also 'forced' and influenced by the digitisation in the society in general to use digital technologies, often without carefully considering the appropriateness and effect of these technologies in context, i.e., consequences on student privacy in K-12 school contexts.

The second problem is that the collected data is largely seen as neutral, pre-analytic and representing an objective state of affairs. Student data is collected, often without sufficient consideration for issues surrounding privacy, security, confidentiality and downstream use of the data. Students' records consisting of demographic, behavioural, learning and relational data are stored, often combined with other information sources and may accompany children across their enrolment, often inter-school. These digital 'dossiers', their scope, and the understandings that inform the data and interventions based on these 'dossiers', become data-doubles that can impact on students' lives, far beyond the context where the data were initially collected, for example by driving decisions that might exclude students from opportunities or curriculum content (Williamson, 2019). In K-12 school settings, where young students have limited data agency, these dossiers have critical implications for their options later in their lives.

The third problem refers to the reality that increasingly, these student data sets are shared, and even sold to commercial and third-party interests as part of service level agreements, data breaches and other practices that violate students' right to privacy, with possible long-term impacts (Regan & Jesse, 2019). If researchers and K-12 institutions continue to develop data analytics projects and infrastructures in order to improve teaching and learning, the obligation to do so responsibly will increase as well. In achieving this, we need to carefully consider different stakeholders' perspectives and practices, as well as to involve them in the co-design of practical ways to be implemented in schools to protect children's privacy in increasingly growing online educational settings. Some scholars, for example, suggest that possible practical ways of

enhancing transparency for children (in regard to how their data is being used) should include “legal visualisation, co-design, co creation techniques and participatory design methods which focus on presenting legal information in a transparent and clear manner” (Milkaite & Lievens, 2020, p.5). Yet, such related practical efforts have hitherto been very scarce. In summary, there is a need for more actionable practice and empirical research that: 1) carefully considers the appropriateness and effect of digital technologies in context, i.e., the consequences on student privacy in K-12 school context; 2) considers where the student dossiers are being used beyond the context where the data was initially collected; and 3) takes into account different stakeholders’ perspectives and practices, as well as involves them into the co-design of practical ways to be implemented in schools to protect children’s privacy.

## 4.2. Data in Use for Teaching and Learning

Novel data analysis technologies and data-based tools, such as Learning Analytics (LA) are increasingly entering the education sector, generating various data on teaching and learning processes and outcomes (Crompton et al., 2022; Holmes, 2022; Vezzoli et al., 2020). How such ‘digital trace data’ (Eynon, 2022) are used in K-12 education and the requirements and consequences of their use are at the core of this report. There are many aspects to data in use for teaching and learning, and in this report, we will highlight a few that we believe are important based on the literature we have covered. This area of research refers to what some term as ‘digital didactics’ (Jahnke & Norberg, 2013), with a focus on the practices of teaching and learning using digital tools and resources.

Of the 38 empirical studies we reviewed on the topic for this section, a majority of them referred to Learning Analytics. However, many of them only indirectly studied learning analytics and rather used the term to focus on topics such as ‘educational data mining for student performance prediction’ (bin Roslan, & Chen, 2022), ‘Data capture and multimodal learning analytics focused on engagement with a new wearable IoT approach’ (Camacho et al., 2020) and ‘Teaching with learning analytics: how to connect computer-based assessment data with classroom instruction’ (Admiraal et al., 2020). This shows the diversity and variety of topics covered by the perspective of learning analytics. Other topics were covered in some articles that we will discuss below.

### 4.2.1. Learning analytics

As mentioned in section 2, Learning Analytics (LA) is a concept frequently mentioned in relation to the datafication of education (Sefton-Green & Pangrazio, 2022). It has gained both much attention and initiated strong debates, especially within higher education, but also in K-12 education. LA in teaching and learning involves the collection, measurement, analysis and use of student data to enhance teaching and learning. LA uses statistics and data visualisation tools to analyse student data and performance from various educational technology systems. Its attention is on using data to develop more accurate predictions and insights from data to enhance teaching and learning

(See for example: Brun et al., 2019 on the METAL project on LA in secondary schools in a French context.) Although LA is important in the context of this report, we are not suggesting that understanding learning in a digital school system can be entirely evidence-based. However, LA can be used for students to develop self-regulated learning or for teachers to make data-informed pedagogical decisions.

Within the research literature we have found several key areas related to LA's wide potential as well as the range of challenges it may bring. First, the potential benefits of LA are wide-ranging and in the reviewed literature we see how it can be used to personalise instruction and experiences of every student by providing differentiated learning plans adapted to student's various needs. Personalised learning can increase student engagement and overall achievements (García-Senín et al., 2022; Hebbecker et al., 2022). LA can also be used as predictive analytics to analyse data on student performance, predict results and student preferences (Tempelaar et al. 2014) as well as to identify at-risk students (Ramaswami et al. 2022).

LA is also being used to identify potential problems enabling timely intervention and providing personalised recommendations for learning materials, learning activities, and/or assessment that is adapted to student's needs and abilities. In a review study on learning analytics literature from 2011 to 2020, Baek and Aguilar (2022) found that 33% of articles they retrieved focused on students with disabilities. Articles that focused on students with disabilities revealed that learning analytics could be used to evaluate and optimise the design of educational technologies, and function as a tool to enhance learning experiences for students with various disabilities. Also, LA can improve teaching practices by monitoring student progress and learning outcomes. In that way LA can help teachers identify effective teaching practices, supporting adaptive teaching methods and improvements in teaching (Schildkamp & Datnow 2022). Data-driven decisions about teaching and learning strategies can also be based on LA.

However, LA also poses significant challenges for students and teachers. Some of the most prevailing ones from the research literature are security, privacy and ethics. There are concerns about privacy when collecting vast amounts of data about students, since protecting data and the privacy of students is essential. Also, obtaining consent from students (through chunks) or other participants is crucial (Salas-Pilco 2022). Furthermore, Schildkamp and Datnow (2022) address the importance of trust in their study from the Netherlands and USA when learning from less successful data use efforts. Yet another challenge is related to teacher education and teacher training and the need to develop digital competence and data-literacy of both pre-service and in-service teachers.

Many educators may not be familiar with LA and its potential to inform practices in the classroom. Also of importance, LA requires accurate and quality data to hinder bias and any form of discrimination. LA is only as unbiased or fair as the data used. Related to this one can mention the platformisation and how big actors such as Google shape the school system by extracting a great amount of digital data (Perrotta et al. 2020; Kerssens, & van Dijck, 2021; Cone, 2023). Lastly, implementing LA in education requires ensuring the necessary resources such as technology

infrastructure, data storage, analytical tools, and human resources, which can be particularly challenging for schools located in disadvantaged areas or which are underfinanced. In the study of Schildkamp and Datnow (2022) we see that for example school organisation and commitment to using data was a hindrance for optimal function of data teams in school.

#### 4.2.2. The Importance of Data Literacy and Data Agency

While European research acknowledges that there are new opportunities for teaching and learning with data, it also concludes that these opportunities cannot be realised without adequate attention to teachers' and students' data literacy (Wasson et al., 2016). Data in use for teaching and learning requires new digital competences of both teachers and students. Teachers need to be supported in their professional learning about using data for the enhancement of their teaching and students' learning. Likewise, students need to be supported in their data literacy and use (Wasson & Kirschner, 2020). (See also the European framework DigComp 2.2.) It is not enough for students to gain functional skills to consider what personal data they provide, manage their privacy settings or respond to data protection options provided by online services. To enact their rights as agents and citizens in a complex datafied society, they also need some measure of critical understanding of the networked data economy which is fuelled not only by data given, but also data traces and inferred or profiled data (Livingstone et al., 2020, p. 232).

According to Wasson et al. (2016) data literacy entails teachers and students developing knowledge, skills and abilities on (1) how the configuration of technology tools or applications impact the data generated (conceptually and technically); (2) what and how data is generated by a tool or application; (3) how the data is analysed (if it is done automatically); (4) how the data/data interpretations can be used in a pedagogical manner (for both teaching and formative assessment); and (5) how the data and data interpretations can be shared (pp. 3). These understandings about data literacy are echoed by Wall et al. (2013) who also bring attention to considering epistemological assumptions associated with data use and how these fit with the practicalities of education in practice. The research by Leeuwen, et al. (2021) in K-12 education in Europe underscores teachers' ability to transform data into actionable instructional knowledge and practices as an important element of data literacy with implications for how novel opportunities for teaching and learning with data are realised. These findings point out the need for more research knowledge on how teachers interpret data available to them and translate this information into action (Leeuwen, et al., 2021; Wasson & Kirschner, 2020).

While recognizing the importance of data literacy, research in Europe equally underscores data agency, that is, teachers' and students' ability and volition to be agentive and contributing members in the data-driven society (Vartiainen, et al., 2021; Wasson & Kirschner, 2020). Data agency is suggested to be enhanced through efforts that empower teachers and students as designers of learning and that support their reflection on their own practice as part of learning and as a way to develop self-regulated learning (Molenaar, et al., 2020; Vartiainen, et al., 2021; Garcia-Senin, et al., 2022). As a response, in Europe attention is given to efforts to support teachers' and

students' data agency through design-oriented and culturally responsive education. It is argued that education needs to move beyond rule-based programming and instructional design, and instead focus on creating opportunities for students and teachers to turn into innovators and designers of learning (Wasson & Kirschner, 2020). Recognizing and cultivating data agency that builds on students' prior knowledge and interests, hands-on learning, embodied interaction and active agency are regarded as prerequisites for equitable and inclusive education. The research by Vartiainen, et al. (2021) shows how engaging in co-design can support students' data agency with enhanced conceptual understanding of machine principles, its workflows, and its role in their everyday practices. At the same time, this research acknowledges that there is still a need for further pedagogical development for students' data agency as it relates to supporting students developing a critical stance toward data and data-driven practices.

Wasson and Kirschner (2020) also point out the need for giving more attention to supporting teachers' data agencies as designers of learning (Wasson & Kirschner, 2020). Here, careful attention needs to be given to considering how increased use of data can support learning design that is culturally and ethically responsive and inclusive. For this, European research calls for participatory methods and cross-boundary collaboration for the advancement of teachers' professional development in data literacy and data agency (Vartiainen et al., 2021). Overall, research on data in use for teaching and learning is emerging along with the tools and technologies, and hence more research in the field is required for understanding the possibilities and consequences of the availability of data to enhance learning and teaching in inclusive, ethical and sustainable ways.

#### 4.2.3. Teacher's professional development

Teachers are key actors in making data use "work". At the same time, the understanding of data and data-based best practices is still very poorly mastered by many teachers (Lockton et al. 2020). Consequently, much of the prior research on data use focuses on how to improve teachers' competencies for using data in education (Kerr et al., 2006; Lockton et al., 2018; Schildkamp & Kuiper, 2010; Vanlommel et al., 2017). However, the use of data to inform practice and new data practices are, also, increasingly seen to be adding to teachers' workload (Selwyn et al., 2023). Below we present the synthesised findings on teacher professional development for data-use in education.

We address the nature of teachers' interactions and agency with data. In the study by Lockton et al. (2020) it is pointed to the fact that the accountability-driven structures and cultures of data use within schools inadvertently constrained teachers' efforts to improve teaching and learning processes. While mandated assessments did not drive their practice, teachers exercised their data agency in for instance using other formative assessment data gathered in their classrooms to guide their instruction either individually or informally with colleagues (e.g. Wasson & Kirschner, 2020). However, according to the authors, they rarely engaged in this activity during their formal teacher collaboration time. Their findings suggest that data need to be especially nuanced to help teachers

clearly identify students' strengths and weaknesses. This reflects arguments from Williamson's (2019) study reporting that the "data-doubles" (how copies of data are travelling) are still far beyond the educational context where the data were initially collected. Efforts to build reflective processes into teachers' lesson plans are well received and help teachers' efforts to use data to further improve their practices (Lockton et al., 2020; Wayman et al., 2017). They also possessed a strong desire to discuss what they were learning about their students with colleagues and to gain insights from one another (cf. Schildkamp & Poortman, 2015).

Data-based pedagogical decision making by the teachers can lead to increased student achievement (see e.g., García-Senín et al., 2022; Hebbecker et al., 2022). At the same time, schools struggle with the implementation of data-based decision making (see e.g., Ebbeler et al., 2016). A key dimension of this issue is teachers' limited data literacy and understanding of how to implement data objects, outputs, analytics, etc. (Leeuwen, et al., 2021; Reeves & Honig, 2015). However, there is also the awareness of an active and agentic role of teachers as change agents (Priestley, Edwards, Priestley & Miller, 2012), as presented above. Professional development in the use of data with a strong focus on teachers' agency is therefore urgently needed, as it can build capacity and engage teachers in data practices and use. In this respect, Schildkamp et al. (2016) developed the professional development program "the data team procedure". They define data teams as professional learning communities engaging in collaborative inquiry, focused on data use. Data teams consist of teachers and school leaders who analyse and use data collaboratively to improve their educational practice and outcomes (Schildkamp et al., 2016). Based on the synthesised findings it seems that the implementation of data teams is a promising way of enhancing the effectiveness of data-based decision making in schools (e.g., Wayman et al., 2017). Whereas Schildkamp and Datnow (2022) study looks at when data-teams struggle. This type of approach is participatory and collaborative, which can also build up teachers' data literacy (e.g., Vartiainen et al., 2021).

Finally, data literacy and teacher's professional development can also be related to teachers' decision processes (Vanlommel et al. 2017), depending on whether the decision process rely on intuitive or analytical processes (Schildkamp et al., 2013). In a meta-analysis, Jung et al. (2018) found that math, reading, writing, and spelling improved when data-based decision-making interventions were implemented. Despite evidence suggesting that data-based decision making can improve academic performance, there is a need for teacher training over time so teachers can make decisions based upon data (Hoogland et al., 2016; Selwyn, 2023). The competence to make informed decisions based on the correct reading and interpretation of student data can ensure that students at the greatest risk are provided with appropriate interventions (Oslund et al., 2021). However, according to Mandinach and Gummer (2016), isolated or "single-shot" professional development sessions are inadequate for the competencies needed to successfully implement data-based decision making. To illustrate, Koh (2011) found that it took two years of ongoing training to improve teachers' graph literacy. Clearly, teachers struggle with reading and interpreting for instance graphs of student data, which requires schools to develop a

comprehensive and recurring professional development plan with respect to the use of data in schools.

Alongside developing teachers' understanding of data, there are also growing pressures of 'data work' practices, which have become part of school leaders, administrators and teachers' work (e.g. Howard et al., 2022; Selwyn, 2020). This is not only about data literacy and working with data, but understanding the ways data is used to measure and evaluate school, teacher and student performance. Howard et al. (2022) have proposed a 'data journeys' approach to make data practices and data use visible to teachers and school leaders, as a way to understand and engage with these practices. A 'data journey' is how, in the case of education, data is selected, transformed and used to evaluate dimensions of education, such as school improvement, teaching practice and student learning. Importantly, teachers and school leaders' understanding and value of data use and related practices will affect interpretation of results, implications for practice and how they understand student progression (e.g. Mills et al., 2021). Data journeys, as a professional learning approach, can make visible data, its movement and associated practices. At this point, data and how it is used can then be discussed, which can support development of common understanding and reflected as part of growing school and classroom data work.

Based on our review of existing research, it can be concluded that while there are widespread discourses around the use of data in teaching and learning, for example, in the context of school effectiveness (e.g., Hopkins et al., 2014; Reynolds et al., 2014) and school leadership (e.g., Bowers et al., 2014; Howard et al., 2022; Schildkamp et al., 2019), less consideration is currently given to addressing the influence of data on the actual pedagogical practices in schools (Krein & Schiefner-Rohs, 2021). Moreover, few studies refer to learning outcomes by students when learning analytics or other perspectives are applied, but rather refer to general impressions of engagement and motivation, or a focus on drill-and-skill learning (Kurcikova, 2022). This is a challenge of the research on data in use for teaching and learning. There is a need for more specific and focused studies documenting teaching, assessment and learning processes and outcomes.

We should not underestimate the complexities involved in ways of handling data by teachers and students. Several studies refer to those working with data in practical settings addressing concerns about the complexities of education data and the "difficulties of ensuring data underpin rather than undermine children's needs and rights." (Livingstone & Pothong, 2022, p. 23), and how this is part of a complex data ecology. Embedded in this is also an awareness of how education data may enable discrimination, exclusion, bias and inequality including gender, ethnicity, sexuality, disability and more.

### | 4.3. Data Governance

This section summarises recent research, 18 studies from our systematic search and additional literature identified by our expertise, exploring questions about data, governance and education. This work largely falls into two camps. First, recent research informed by sociology and critical

policy studies has explored how data can be used as a form of knowledge for governing education systems, schools or even individuals. Here, governance is understood in terms of the institutions, technologies and practices that control, steer and shape actions in education, ranging from the official policies that govern whole systems and schools to the daily school policies, practices and technologies that govern teachers' work and students' learning. These works, collectively, focus on governance through data. Second is a body of work focusing specifically on data governance, or the governance of data in education systems. This work, informed more by legal, regulatory and rights-based perspectives, as well as theory of organisations and political science, seeks to explore how educational data are managed and protected, or exploited and mis-used, and is aimed at identifying models of good governance of school data. (See also section 4.1.) We start by summarising research on governance through data, followed by the research on governance of data.

#### 4.3.1. Education governance through data

Education governance refers to the institutions and practices through which schools are managed, regulated and controlled, usually understood to be synonymous with education policy mandated by state agencies of government. However, in recent decades a shift has taken place across European education systems. Centralised governmental control has become more dispersed and decentralised, with other non-state actors such as consultancies, commercial companies, charitable organisations and other public-private hybrids taking up important governance roles (Ball, 2021). A 'global education industry' has emerged, with a vast range of public and private sector organisations participating in various aspects of educational service provision and delivery (Verger et al., 2016).

One key function of the global education industry has been to advance the uses of digital data for purposes of educational governance. The use of numerical information for measuring various aspects of school or system performance has a long history going back to the nineteenth century, but has become particularly significant in the twenty-first century as a means of monitoring and improving teaching, learning, and school management (Lawn, 2013). The turn to data over the last two decades has been advanced by the rising demands of performance-based accountability, and the collection and use of data for performance monitoring purposes has become widespread in European schooling systems as well as beyond (Piattoeva and Boden, 2020).

Central to such efforts has been the construction of complex and sophisticated computerised systems that can collect, store and process very large quantities of data generated through examinations, tests, and various forms of school inspection and auditing, often resulting in particular evaluations and judgments of school performance (Grek et al. 2021; Ozga et al., 2011; Verger et al. 2019). These systems are called data infrastructures, and range from those built to enable the international tests conducted by organisations like the OECD or national state assessments (Hartong, 2018), to commercial testing systems routinely used by schools for monitoring student performance in a range of subject areas (especially literacy and numeracy) and

various competencies and skills (like problem solving, or social and emotional skills) (Williamson & Piattoeva, 2019).

Vast data infrastructures have therefore become highly significant in various forms of school monitoring, measurement, inspection and accountability practices in Europe and beyond (Hartong & Förschler, 2020). While numerical data infrastructures have long been integral to the governance of education, other forms of data have also become significant in governing the schooling sector, particularly 'big data' generated from the everyday software systems used in schools across Europe and internationally. However, there are not many empirical studies on these practices yet.

#### 4.3.2. Data-driven digital governance

Internationally, schools are said to be increasingly 'digitalized' and 'datafied', with key functions and processes performed by software packages and infrastructures that also routinely gather data about institutions, staff and students (Pangrazio et al., 2023). The digital transformation of education in recent years means that school data, especially students' data, are now increasingly collected by education technology businesses (the EdTech industry) and even multinational technology corporations (Big Tech) that provide educational services (Day et al, 2022). Both EdTech and Big Tech industry operators are becoming influential actors in education systems, using their existing technical and market power to extend new digital services into schools. The result is that technology companies, ranging from education-specific providers to global corporations like Google, Microsoft and Amazon, are now playing significant roles in the governance of education, by governing and shaping how pedagogy, curriculum and assessment can take place (Williamson, 2023). For example, new digital applications widely used in schools can affect how teachers plan and conduct lessons, how they source and deliver curriculum content, and how assessment data are generated and used.

Digital data-processing technologies perform actions that constitute emerging forms of 'digital education governance', whereby governance is not only enacted through official policy prescriptions or mandates, but software affects how schools, teachers and students do things on an ongoing basis (Williamson, 2016a; Landri, 2018). Digital governance of education is based on the widespread collection of 'digital trace data' generated from students' interactions with computer software (such as a learning management system, or a learning platform) (Eynon, 2022). Digital data processing technologies make it possible to provide fine-grained assessments and analyses of individuals in or near 'real time' as they interact with an interface and enable data and feedback to flow synchronously and recursively within the pedagogic apparatus of the classroom itself (Williamson, 2016b). This data speeds up the flow of information about students, affecting how teachers and leaders can know and relate to students—as digital data profiles—and informing pedagogic decision-making and other institutional procedures. These processes are exemplified by the trend towards data-driven 'personalised learning', where software can prescribe an individual's course of study based on predictions derived from past data about their progress, and by forms of

data-driven decision making in schools, where choices about the allocation of resources or the targeting of support are driven by an institution's aggregated data (Perrotta and Selwyn, 2019).

The implications for how education is governed are significant, as industrial data-processing technologies play a part in reshaping many routine aspects of schooling. This means that the organisations and companies that build digital technologies for education are influencing practices in classrooms and school administration. Their influence plays out on a daily basis, in contrast to the bureaucratic pace of official policy implementation. They constitute a largely invisible digital architecture of education governance, which does its own kind of policy work by setting coded instructions and algorithmic rules for how things can or should (and cannot or should not) be done in schools. In this sense, schools are more and more 'governed by EdTech' (Kerssens & van Dijck, 2022).

### 4.3.3. Platform governance and educational autonomy

One of the main ways that digital education governance occurs in schools is through the expanding range of digital education platforms available to schools. The term 'platform' refers to a product that is internet-based (rather than being a standalone software package), can be modified and updated by its producers, enables engagement from users, and continuously collects data from users (van Dijck et al., 2018). Understood technically, educational platforms are interfaces through which students, teachers and leaders can access content and functionality—but they can also exert profound effects through processes known as 'platform governance', where private platforms are positioned as technologies to achieve public policy aims:

Not only does platformization cede tremendous influence over the everyday activities of public education systems to private technology companies, but it also tethers institutional decision-making to the governing logics of the code, algorithms, and data processes that underwrite platform technologies. (Nichols and Dixon-Roman, 2024, pp. 309-310)

Digital platforms can reshape the forms that teaching and learning take, bring commercial companies directly into the classroom, and accelerate and amplify the collection of student data (Decuypere et al., 2021). Such platforms range from those that focus on behaviour monitoring, such as the widely used ClassDojo, to school-wide learning systems like Google Classroom (Hooper et al, 2022). Through a marketplace of digital education platforms that has expanded in Europe and beyond in recent years, public and state schooling systems are increasingly underpinned by private technologies, bringing about concerns over the erosion of public values by private interests and profit-seeking motives (Kerssens and van Dijck, 2021). In other words, platform governance also enables processes of privatisation and commercialisation to take place, whereby private technology platforms can easily become embedded in public education systems, mine new sources of digital data, and reshape what happens in schools (Cone et al, 2022).

The governance of schools by EdTech platforms also has significantly wider implications for schools' autonomy and teacher professionalism (Dabisch, 2023; Kerssens & van Dijck, 2022).

Platforms may be reconfiguring teachers' professional autonomy in two key ways. On one hand, the introduction of platforms into schools is accompanied by calls for teachers to be equipped with new skills and literacies to be able to incorporate platforms into their practices, and to make use of the data produced by platforms to inform their professional knowledge, planning and assessment practice. On the other hand, platforms can be understood as eroding teachers' professional autonomy and leading to de-professionalization by 'taking professional autonomy and judgement away from educators, whilst at the same time empowering technology providers and algorithmic systems of decision-making to increasingly influence what is happening in various educational practices' (Hartong & Decuyper, 2023, p. 2). Specific forms of platform governance of schools, then, bring a particular commercial logic into educational environments and practices, and make schools dependent upon EdTech platforms for many everyday processes.

#### 4.3.4. Governing stacks and creeping commercialization

Education platforms are underpinned by much bigger and far-reaching technical systems that are sometimes called 'stacks' (Nichols & Garcia, 2022). Stacks refer to the complex networks of technologies that a platform depends on to function, and includes things like data storage, data analytics, and cloud computing power (van Dijck, 2020). These stacks are largely invisible to schools, but are significant because they are owned by some of the biggest global technology firms, including the Big Tech operators Microsoft, Google and Amazon, giving them a widespread presence in schools (Perrotta et al, 2020; Williamson et al, 2022). Google and Microsoft are present in many schools both as providers of education platforms—like Google Workspace, or MS education tools—and this already connects schools to their global stacks of cloud computing and data facilities, while Amazon reaches into school through a different route by providing cloud and data functionality for third party EdTech vendors (Williamson, 2023).

When a school makes use of an educational platform, it is also often, but not always, connected to a Big Tech stack that can itself exert significant governing effects on schools' everyday operations. Veale (2022, p. 76) argues schools are slowly becoming 'extremely reliant on a few large companies' entire technological stacks in order to operate', which are 'reshaping what schooling is and could be, and exercising unaccountable control over students, teachers, administrators and content providers alike'. Further, this control over education also involves the widespread extraction of student trace data, which can then be used for developing 'artificial intelligence' (AI) applications such as personalised or adaptive learning – an aim widely shared by those Big Tech operators with interests in education:

Insofar as deployment of AI in education is broadly yet-to-come, it is crucial that schools and related decision-makers grapple now with the political economy of the technology stacks they are enmeshed in, in order not to lose further control of key pedagogical choices in the years to come. (Veale, 2022, p. 77)

The development of AI for education, then, may advance school dependencies on Big Tech and entrench technological lock-ins while reshaping pedagogic practice (Williamson et al, 2024). Such lock-ins advance corporate control over the technical foundations of teaching, learning and school management, empowering private technology companies to claim control over student data as a route to generating value. On this point, van Dijck (2020, p. 2808) notes that 'the dependence of schools on proprietary information systems effectively funnels pupils' data, generated in a public context, into a proprietary data flow controlled by one corporation's platforms'. Notably, governments in some European countries and regions have begun pushing back against Big Tech expansion of state education, indicating rising concern about the power of these companies to govern everyday school operations and extract student data at scale (Hooper et al., 2022).

#### 4.3.5. Automated governance and pre-emptive judgement

One of the main forms of governance enabled by platforms and the technology stacks underpinning them can be understood as automated and anticipatory governance. Witzemberger and Gulson (2021) refer to 'automated education governance' as 'system-level and school-level practices that are exercised by automated decision-making machines, and instances in which software has a role in governing education'. Automated governance operates through AI such as machine learning, which uses data to model and predict school performance and students' cognition, knowledge, comprehension, and even mental health, and is based on the assumption that education can be augmented, enhanced and improved by data analytics, machine learning and automation (Selwyn et al, 2023). The technologies and data practices underpinning automated and anticipatory governance include accelerated 'real-time' measurement rather than reliance on discrete assessments and evaluations conducted at long temporal intervals; predictive analytics that make probabilistic calculations of a student's future progress; automated responses, feedback, and 'personalised learning' adaptations; and even 'pre-emptive' interventions generated by AI itself (Witzemberger & Gulson, 2021).

Automated technologies are significant for educational governance because they can automate decisions about students by processing student data, making predictions about their likely progress, and automatically producing feedback or 'adaptive' interventions (Hansen & Komljenovic, 2023). This is sometimes referred to as both predictive and prescriptive learning analytics, and has been a key aim of the education technology industry under the banner of personalised learning (Williamson, 2016c). The key issue here is that the use of such technologies may lead to an increase in 'automated judgement' and a reduction in professional judgement by teachers, with potentially high-stakes decisions that could affect students' future educational pathways offloaded to external systems (Selwyn et al, 2023).

While automated forms of judgement may be attractive in educational systems where efficiency and time savings are sought after, automation may also make errors that are inexplicable and hard to trace (Bergviken et al., 2023). There are also significant questions about the potential for bias and discrimination in such judgements. Automated and data-driven decision making might

exclude certain groups of students from access to particular knowledge or opportunities on the basis of historical data indicating such groups are less likely to perform highly on academic assessments, and thus reproduce existing inequalities (Hakimi et al., 2020). But these problems and errors may go unnoticed due to the very design of the software. With proprietary EdTech or Big Tech platforms and stacks, it is often impossible to ascertain how automated decisions are made as the internal mechanisms leading to particular forms of judgement or discrimination are protected by confidentiality and intellectual property laws.

#### 4.3.6. Data ownership

A significant issue with many examples of governance through data is that the data fall under the ownership of commercial entities, for whom the data afford significant monetary value (Zuboff, 2015). This means that while an educational platform may pose risks in terms of student opportunities, it carries financial potential for the platform proprietor, which can leverage large quantities of student information as the basis for producing new derivative products or upgrades (Komljenovic, 2021). Digital data are thus the basis of a pervasive business model in the educational technology industry, as well as for Big Tech operators like Google, Amazon and Microsoft, on whom many third party EdTech firms themselves depend for computing power, data storage and analytics capacity (Williamson, 2023). This is a business model that extracts two kinds of 'rent' from educational institutions. The first kind of rent is monetary, as schools pay fees or subscriptions to use a platform or service; and the second is 'data rent' as schools ultimately provide value to platform proprietors or service operators by allowing them to extract digital information about students, staff, and institutional processes (Komljenovic, 2021).

By extracting data rent from schools, companies can therefore claim ownership over schools' data and exploit it for value-creating ends. Thus, for platform owners, the displacement of teachers' professional autonomy and judgement by an automated function is a value-creating opportunity, as the processes, places, and authorities of education judgement are moved away from the teacher towards platforms for which schools pay a subscription fee (Hansen & Komljenovic, 2023).

Financial investors are enthusiastic about this approach as it appears to promise perpetual income generation, as platform operators continually upgrade their services using data to inform product development and command the ongoing extraction of subscription payments from schools (Davies et al, 2022).

Educational data ownership is therefore consequential because it may empower education platform proprietors to redefine some of the key ways that pedagogy is enacted and curriculum materials are accessed, according to calculations made about the potential monetary value to be gained. In the process, these economic calculations result in the increasingly pervasive use of data-driven technologies in schools and classrooms, in ways that have governing effects on pedagogy, curriculum and assessment practices. The business model of data-driven educational technology in this sense is inseparable from the forms of educational governance and control practised by both EdTech and Big Tech firms. How these forms of governance of education through data actually

affect practices within schools remains under-researched, and there remains a clear need for further studies that analyse the technical and business models of education technology providers, and their contextually situated effects in different schools and education systems (Decuyper et al, 2021; Pangrazio & Sefton-Green, 2022).

#### 4.3.7. Governance of data

The ownership of data and responsibility for data also raise issues of what good data governance and regulation should look like when either state or commercial entities are collecting it in 'datafied schools' (Henne & Gstrein, 2023). In the UK, for example, repeated questions have been raised about the Department for Education's governance of the data in the National Pupil Database, one of the largest repositories of student information in the world. The DfE makes the data available for analysis by third parties, including for commercial re-use, often without fully transparent purposes, which has led to significant instances of misuse of student data (Persson, 2022). Ultimately, this is a problem of governance at the Parliamentary level: 'Although good governance and the rule of law require primary legislation to be laid down by Parliament and to be subject to debate, the UK government has not produced legislation specifically concerning EdTech or education data' (Day et al., 2022, p. 2).

When it comes to commercial companies providing educational platforms and underlying stack services, data governance models also raise important questions about appropriate management and safeguarding of student information. A UK-based legal study found that some of the most popular classroom platforms have opaque data protection and privacy policies and legal terms that make it difficult to ascertain what data they collect, how they use it, or how it is shared, and which are 'inconsistent with data protection laws and could result in the commercial exploitation of children's personal data' (Hooper et al., 2022, p. 7). In 2022 the US Federal Trade Commission (FTC) warned the EdTech industry that it would be enforcing sanctions against companies collecting more data than absolutely necessary for their service to function, while both US lawmakers and European regulators have begun legal challenges against EdTech providers such as Google (Hooper et al., 2022).

In this context, schools are often positioned as data controllers with responsibility for governance of students' data. But 'the corporate power of EdTech, its ethos of data maximisation (rather than minimisation), and commercially motivated policies and designs place a near-impossible burden on any school, parent, caregiver or child wanting to manage how data processed from children are used' (Hooper et al., 2022, p. 9). Data maximisation refers here to an EdTech provider collecting as much user information as possible, often for purposes that extend far beyond what is necessary to run the service, with various groups increasingly calling for EdTech to minimise the data collects for more tightly defined purposes. As Day et al. (2022) have outlined in the UK context, current legal and regulatory instruments fail to contain the collection of student data by EdTech, leading to schools' poor compliance with data protection laws:

Not only do schools lack expertise and resources, as well as straightforward enforceable guidance from the government, but there is a considerable power imbalance between schools and EdTech providers. This avails EdTech providers instant access to data collected from children in schools with insufficient oversight, transparency or accountability and with considerable latitude to interpret and apply the law as they choose. (Day et al, 2022, 10)

Consequently, there are rising calls across Europe for forms of data governance that might better support responsible, inclusive and meaningful use of data. To address this issue of data governance, the Digital Futures Commission (DFC) in the UK has recently released a blueprint document setting out a baseline for data processing, which businesses and schools should follow (Digital Futures Commission, 2023). The governance framework it offers is based on existing international child rights instruments and is aimed at protecting children from commercial exploitation by education technology companies, and ensuring compliance with safety, security, privacy and data protection standards, using certification to incentivise compliance with these standards across the EdTech sector. The three central priorities for good education data governance it provides are:

1. Clarifying, and where necessary, extending the relevance of existing frameworks that protect children's data to ensure a coherent regulatory environment. This includes routinely upholding the UN Convention on the Rights of the Child; ensuring compliance with GDPR particularly around transparency of data processing, purpose limitation, data minimization, and integrity and confidentiality; and enforcing the Age-Appropriate Design Code, introduced in the UK in 2018, with which many education technology companies currently fail to comply.
2. Introducing certification to ensure compliance and measure learning benefits for EdTech used in school settings. Modelled on good practice in health and social care, this would involve creating a standardized comprehensive framework for assessing and approving EdTech regarding its educational values and benefits, opportunity costs or risks, usability and accessibility, interoperability, data protection and privacy, and security. A key aspect of this proposal would include more meaningful oversight of the types of data collected about children (factual personal data and inferred or behavioural judgments), how they are processed, and transparency about the construction of inferred data.
3. Developing trusted data infrastructure(s) for research, business and government that serve the best interests of children and the wider educational community. While commercial companies are able to harvest student data at scale for commercial purposes, these data are usually unavailable to researchers and/or other civic society and public agencies with responsibility for children's education due to intellectual property or commercial reasons. The DFC therefore suggests the development of an operational model of education data sharing, to include data processed by EdTech, in the public interest, including a clear framework for data access and an easily accessible system of redress for children whose data have been exploited.

The DFC proposal, while primarily targeted at UK bodies, could provide the basis for Europe-wide public and policy debate regarding good governance to support responsible, inclusive and meaningful use of educational data. However, there are also several emerging initiatives for comparison of generative AI governance frameworks in different education systems, for example by the OECD (2023), that will be important in the years to come.

# 5. Towards an Analytical Framework

**The rationale:** There are two key perspectives creating the rationale and outline for this report. The first perspective is addressing the ecosystem of datafication in and of education, which implies a need to study the system as a whole and how parts in the system are interrelated. As part of the Agile-EDU project the concept of ecosystem refers to the definition made by Diaz-Gibson et al. as 'social infrastructure formed by diverse actors that share a purpose, and engage in collaboration to co-design and co-implement innovative responses to existing social and educational challenges.' (referred to in Global Education Futures 2020). The ecosystem also raises awareness on how we understand the complexities involved in datafication on different levels. Even when we are focusing on one topic, we are aware of the interconnections to other parts of the system. The second perspective is equally important since it is about the fundamental values of our education system, especially on inclusion and equity. Education, as an ideal in a European context, is about creating equal possibilities for all within democratic societies. The fundamental question today is how datafication and the role of global BigTech companies threaten and challenge these values and education as an ecosystem. As such, inclusion and equity should be thought of as embedded in every part of our analytical framework, as an issue to consider and implement.

We argue that this ecosystem is based on the interrelationship between different levels of analysis; a) practices in classroom activities, b) regional and local aspects of data governance and c) national and European regulations, rights and privacy legislations and norms, understood as dynamic processes.

**The core:** At the core of the analytical framework are the digital data generated by learners in educational practices, as individual students, but also in interaction with other students, as well as teachers. This is illustrated by the representation of a student working on a digital device together with others in the Figure below. Without any activity using digital devices or registration of student data, there is no "start" of the ecosystem of education data. By providing a model that puts the making of these data by learner activities in the middle, we aim to illustrate the process of collecting, sorting, analysing or commodifying such data as datafication. This core and the process of datafication is framed within specific technological infrastructures referred to as digital

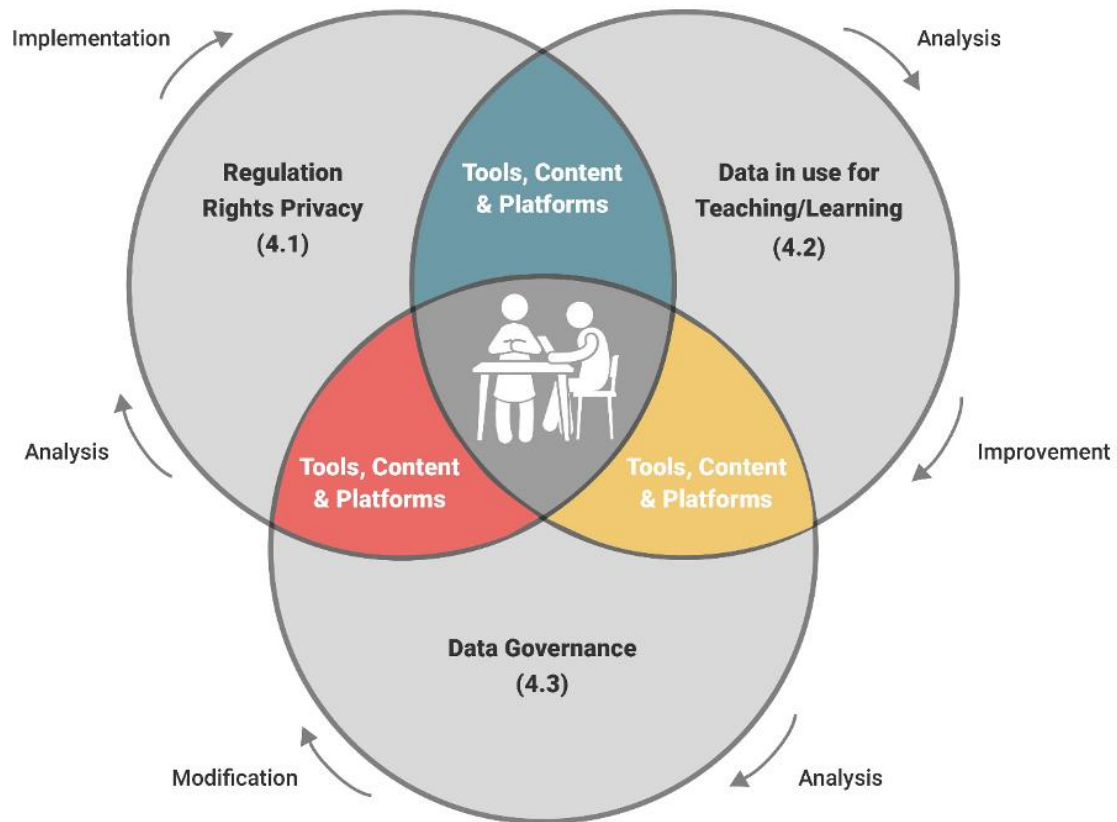
platforms and the platformisation of everyday life and educational practices, which are constantly developed by EdTech/BigTech as providers and stakeholders in the education data ecosystem. In the analytical framework we have referred to this as 'tools, content and platforms' that students interrelate with in ways of creating data. Different stakeholders provide a diverse set of digital tools, contents/resources and new digital infrastructures/platforms, also including learning analytics (LA) and artificial intelligence (AI), that create both new possibilities and new challenges for educational practices. This is also what teachers today have to relate to in their teaching and assessment on different levels of K-12 education.

**The circles:** The three main topics covered in this report are visualized as three circles; 'Regulation, Rights, Privacy' (4.1), 'Data in use for Teaching/Learning' (4.2), and 'Data Governance' (4.3). Each of them has its own dynamics expressed as a constant process of analysis and ways implement, improve and modify aspects of datafication in and of education. These topics highlight some of the critical issues and themes from the research literature. Concerning our guiding question mentioned in the introduction, we have documented what the research literature can inform us about the conditions and criteria for responsible, inclusive and meaningful use of data to support teaching and learning. The aim of the analytical framework is to understand the intersection between these three topics. In order to visualise this, we have chosen a Venn diagram. We are well aware that such a model might not be able to explain every detail in such a complex ecology that school and datafication represent. However, a Venn diagram may present visually three overlapping fields, which might have an analytical interest for the purpose of understanding the connection between the three fields of research we have investigated, of key importance for the Agile EDU project.

The left circle refers to section 4.1 in the report where we showed how data regulation, rights and privacy has been addressed on both European, national and regional levels, but also challenges in ways of handling student data. For 'Regulation, Rights, Privacy' this is expressed as analysis of new technological developments and implementation of new regulations and rights in this area, which needs to be constantly renewed. The discourse of rights and privacy in the European Union, for example, relates to ongoing work on the AI Act, GDPR and other frameworks for data regulation. In order to understand and visualise this discourse, we illustrate it with the left circle in the model, and we will return to how this discourse is related to the actual teaching and learning in school. However, our review identified that there is a need for more actionable practice and empirical research focusing on these topics. The right circle refers to section 4.2, focusing on data in use for teaching and learning on a local level. The dynamics of this circle is related to how teachers and students analyse and use data to improve teaching and learning. Important are students' activities using digital technologies, software and applications, and how data is generated as specific algorithms, either personal or process data. These data may, in some cases, be a national test, digital footprints for the teachers to act upon or valuable patterns for an EdTech company. One click can, in other words, have diverse value for different stakeholders. Data can then be studied as an object in itself or as part of the process of teaching and learning. Regarding the latter, it is important for teachers to work with new competencies, data-literacy, in order to understand data and the process of datafication, including learning about platforms and diverse software that

include learning analytics and artificial intelligence. Teachers, as professional learning communities, including their headmasters and other resources within a smaller municipality or/and region of schools, establish and elaborate local practices in the use of data and learning analytics. These communities of practice, working with data in school, do also influence teaching and learning in schools, but in different ways than the frameworks for data regulations, rights and privacy. This section shows that there is not much empirical based, systematic research on using educational data by teachers to improve educational practices and students' learning. There is a lack of empirical studies that have looked at teachers and their teaching, when they use data and data-driven decision making. The bottom circle refers to the growing importance of data governance, either with a focus on governance through data, understood as a form of knowledge for governing education systems, schools or even individuals, or governance of data, understood as how educational data are managed and protected, or exploited and mis-used, aimed at identifying models of good governance of school data. The dynamics of this circle is about ways of analysing through and of data for modification of education systems, institutions and individuals.

The tensions: We argue that much of the essential discussions around students' data must be carried out as boundary work in shared spaces across the three sections presented in this report. There are embedded tensions, challenges and possibilities within the different shared spaces between the different circles in the figure that needs to be explored further. In order to address these tensions and overlapping areas, we have used different colours (see Figure below). We have formulated these important tensions as ways of handling datafication as related to the developments of 'tools, content and platforms' for each intersection between circles, but defined in different ways. 'Tools, content and platforms' express ways of using data (datafication). The intersection between 4.1 and 4.2 is much about how national and European regulation, rights and privacy issues related to 'tools, content and platforms' are implemented into practices in schools and classrooms. The overlap and tension between 4.1 and 4.3 is mostly about how national and European regulation, rights and privacy issues are impacting on ways of data governance, and vice versa. The tension between 4.2 and 4.3, as stated earlier in this report (section 4.3.6), is about how forms of governance of education through data actually affect practices within schools when using 'tools, content and platforms', and how the technical and business models of education technology providers have an impact on contextually situated schools and education systems. However, data in use, as expressed as data literacy and agency, also has an impact on ways of understanding data governance, also expressing potential tensions.



**Figure 1: Ecosystem of education data for K-12**

Importantly, this dynamic ecosystem is generated towards improving students' learning and teacher's teaching. Recent developments of digital tools, content and platforms, adapted for student's learning in different domains and on different levels, have provided many new and exciting possibilities. At the same time, as we have emphasised in this report, there are important challenges and problematic developments that threaten our education system in ways of handling datafication in and of education.

This literature review has been focused on drawing on existing research in this area and documenting the status of knowledge. And as has been stated several times throughout the report, there is a lack of systematic research in many of the topics covered in this report. There is a message from our work targeted the need for more systematic and evidence-based knowledge of high importance for contemporary European education systems.

# 6. Defining key themes and questions

There is a fundamental tension underlying all of what has been presented in this report, between the possibilities that datafication represents for students' learning processes and outcomes on one side, and the challenges the management of data implies on the other.

Below we present some key themes and question formulations to each theme that should be further discussed, explored and acted upon. These key themes are identified based on the literature review in section 4 above and our analytical framework presented in section 5. Potentially there are a number of themes and questions in this area, but we want to highlight the following:

1. Key theme: A need for a systemic approach (ecosystem) towards datafication.

Key questions: How are the different components of the analytical framework interrelated? Which stakeholders are involved? How does the analytical framework capture different processes in your own country? Whose interests are involved in educational data, and for what purposes?

2. Key theme: New challenges for inclusion and equity are developing, and new digital divides are appearing.

Key questions: How equity is defined in a specific project or educational context? How is access to digital devices and educational data ensured across European countries? What kind of data do students and teachers have access to? Does lack of AI literacy or data literacy create a new digital divide? How can school authorities secure data sharing practices? How can we support student's and teacher's digital competence in using data and understanding data sharing practices?

3. Key theme: We know very little about students' learning, datafication and learning analytics (LA).

Key questions: How can school authorities contribute to improved and responsible student's agency related to LA? How can student's conceptions, understanding of LA and digital technology be strengthened? How are students' digital responsibility and digital citizenship developed? How can digital didactics be developed to support student's learning and teaching practices?

4. Key theme: It is important to focus on tensions between legislation on a European and national level and the challenges of implementing these regulations and rights on a regional and local level.

Key questions: How do we move from European and national regulations and legislations to local implementation in schools and practices? What kind of understanding and guidance do school authorities and school leaders need? What kinds of professional developmental programs are needed?

5. Key theme: The importance of 'multiliteracies' (data literacy, AI literacy, computational thinking, ++).

Key questions: What is data literacy and how can it be taught? How can we address questions concerning gaps in data literacy among students, teachers and school leaders?

6. Key theme: The challenge of developing 'multiliteracies' as part of both in-service and pre-service teacher training.

Key questions: How to make sustainable networks (teacher's teams) for developing data literacy and key competences for in-service teachers on a regional and/or local level? How can national authorities and national guidelines support teacher educators and pre-service teachers?

7. Key theme: Awareness of the importance of ethics and biased data in machine learning needs to be raised and strengthened.

Key questions: How can EdTech address quality and bias of data? How can teachers address the challenges of biased data and work with students on the challenges raised in biased data? How can EdTech design their data-driven products to ensure that students and teachers are empowered in their practices? How can EdTech ensure that their products support the development of all key skills? How can data-driven decision making be reconciled with teacher professional judgements?

8. Key theme: Awareness of education data ecosystems and EdTechs/BigTechs impact on schooling and education.

Key questions: How can policy and practice manage the increased influence of commercial stakeholders within public and private education? Which new possibilities and challenges emerge as new educational stakeholders have impact on education as a system and within classroom practices?

# 7. Education Data Futures

This report has been written with the intention to stimulate discussion and debate, as well as different activities in the Agile-EDU project, about one of the most important issues for contemporary education systems – that is 'datafication in and of education'. Datafication, and different associated issues (platformisation, LA, AI), have emerged as fundamental challenges of our times and for future citizens. Within just a few years new commercial stakeholders have gained influence within education and our everyday lives, often in subtle and diffuse ways that challenge policy and legislation, as well as educational practices. These developments trigger questions about what the future of education might look like.

Of course, technological developments have always embedded conceptions on the future of education. In his book 'Big Data in Education' (2017) Ben Williamson use the term 'sociotechnical imaginaries' (p. 16) from the field of science and technology studies (STS) referring to future conceptions of the interrelationship between technological developments and education. By using this term Williamson refers to a definition by Jasanoff (2015) defining 'sociotechnical imaginaries' as "collectively held, institutionally stabilized, and publicly performed visions of desirable futures, animated by shared understanding of forms of social life and social order attainable through, and supportive of, advances in science and technology." (p. 4) This literature review has tried to explore what this means for the area of datafication as related to 'Regulation, Rights, Privacy', 'Data in use for Teaching/Learning', and 'Data Governance', as ways of imagining education data futures.

Across Europe there are initiatives that explore what these developments might imply for different sectors of society such as education. As shown in different parts of this report several initiatives have evolved in just a few years, such as the implementation of GDPR, the Council of Europe initiatives on AI, regional and national initiatives (for example as in Norway with a governmental committee on learning analytics), and strategic inter-sectorial initiatives within this field, such as the Digital Futures Commission in the UK.

In recent years there has also been a growing awareness of what is termed 'digital disconnection', both as public movement among parents and teachers of restricting the use of digital technologies at home and in schools and as a field of research in media studies (Lomborg & Ytre-Arne, 2021; Bagger, 2024). This research represents a critical response to the mediated conditions that characterize contemporary societies and the impact of digital technologies in everyday life of citizens, and can be defined as a counterforce to the pervasive nature of using technologies and digital data in different social contexts.

A key challenge is the rapid pace of developments of technological innovations. It is difficult to grasp what will be around the next corner. For example, not many had heard about ChatGPT three months ago, and even though artificial intelligence has been with us for a long time, only recently it has been discussed all over in the public discourse. However, this is not only related to rapid pace, but also as a challenge of scale and scope of transformations. Datafication represents fundamental challenges of the traditional educational ecosystem, and there is a growing awareness of the impact of commercial interests. Recently, even directors of big EdTech companies have asked for a time out in developments of AI at this moment in time as shown as part of congress hearings in the US during 2023-2024. Although it might not be such a time out, we need, as educators, to raise awareness about what these developments in AI represent for education. At the same time, education can play a key role in ways of engaging our future citizens in questions about education data futures, as digital competences and 'multiliteracies'. As such, teacher education and in-service training (see e.g., Nazaretsky et al., 2022) will be of great importance in defining future developments and imaginaries in the field of datafication.

Looking ahead it will be important to discuss how we can further develop the education data ecosystem and the different forces influencing it and the ethical dilemmas that might emerge.

# 8. References

- Admiraal, W., Vermeulen, J., & Bulterman-Bos, J. (2020). Teaching with learning analytics: how to connect computer-based assessment data with classroom instruction? *Technology, Pedagogy and Education*, 29(5), 577–591. <https://doi.org/10.1080/1475939X.2020.1825992>
- Baek, C., & Aguilar, S. J. (2022). Past, present, and future directions of learning analytics research for students with disabilities. *Journal of Research on Technology in Education*, ahead-of-print(ahead-of-print), 1–16. <https://doi.org/10.1080/15391523.2022.2067796>
- Bagger, C. (2024). A decade of digital disconnection research in review: Where, what, how, and who? In K. Albris, K. Fast, F. Karlsen, A. Kaun, S. Lomborg, & T. Syvertsen (Eds.), *The digital backlash and the paradoxes of disconnection*, pp. 109–128. Nordicom, University of Gothenburg. <https://doi.org/10.48335/9789188855961-6>
- Baker, R. S., & Hawn, A. (2022). Algorithmic Bias in Education. *International Journal of Artificial Intelligence in Education*, 32(4), 1052–1092. <https://doi.org/10.1007/s40593-021-00285-9>
- Ball, S. 2021. *The Education Debate*, 4th ed. Policy Press.
- Benson, E. (2020). *Surroundings: A history of environments and environmentalism*. University of Chicago Press.
- Bergviken Rensfeldt, A. and Rahm, L. 2023. Automating Teacher Work? A History of the Politics of Automation and Artificial Intelligence in Education. *Postdigital Science and Education*, 5, 25-43.
- Biesta, G. (2009) Good education in an age of measurement: on the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21 (1), 33–46.
- Bill Cope, Mary Kalantzis & Duane Searsmith (2021) Artificial intelligence for education: Knowledge and its assessment in AI-enabled learning ecologies, *Educational Philosophy and Theory*, 53:12, 1229-1245. <https://doi.org/10.1080/00131857-2020-1728732>
- bin Roslan, M. H., & Chen, C. J. (2022). Educational Data Mining for Student Performance Prediction: A Systematic Literature Review (2015-2021). *International Journal of Emerging Technologies in Learning*, 17(5), 147–179. <https://doi.org/10.3991/ijet.v17i05.27685>
- Black, P. & Wiliam, D. (1998) 'Assessment and Classroom Learning', *Assessment in Education: Principles, Policy & Practice*, 5:1, 7 — 74
- Bowers, A. J., Shoho, A. R., and Barnett, B. G. (2014). *Considering Use Data By Sch. Leaders Decis. Making: Introduction*. Information Age Publishing Inc. <https://doi.org/10.7916/D8862F32>
- Brun, A., Bonnin, G., Castagnos, S., Roussanaly, A., & Boyer, A. (2019). Learning analytics

- made in France: the METAL project. *The International Journal of Information and Learning Technology*, 36(4), 299–313. <https://doi.org/10.1108/IJILT-02-2019-0022>
- Bulger, M., McCormick, P., & Pitcan, M. (2017). The Legacy of InBloom. Working paper. Data & Society. [https://datasociety.net/wp-content/uploads/2017/02/InBloom\\_feb\\_2017.pdf](https://datasociety.net/wp-content/uploads/2017/02/InBloom_feb_2017.pdf)
- Camacho, V. L., Guia, E. de la, Olivares, T., Flores, M. J., & Orozco-Barbosa, L. (2020). Data Capture and Multimodal Learning Analytics Focused on Engagement With a New Wearable IoT Approach. *IEEE Transactions on Learning Technologies*, 13(4), 704–717. <https://doi.org/10.1109/TLT.2020.2999787>
- Carvalho, L., Martinez-Maldonado, R., Tsai, Y. S., Markauskaite, L., & De Laat, M. (2022). How can we design for learning in an AI world?. *Computers and Education: Artificial Intelligence*, 3, 100053.
- Cone, L. (2023). The platform classroom: troubling student configurations in a Danish primary school. *Learning, Media and Technology*, 48:1, 52-64.
- Cone, L., Brøgger, K., Berghmans, M., Decuyper, M., Förschler, A., Grimaldi, E., Hartong, S., Hillman, T., Ideland, M., Landri, P., van de Oudeweetering, K., Player-Koro, C., Bergviken Rensfeldt, A., Rönnerberg, L., Taglietti, D., & Vanermen, L. (2022). Pandemic Acceleration: Covid-19 and the emergency digitalization of European education. *European Educational Research Journal*, 21(5), 845–868.
- Crompton, H., Jones, M., & Burke, D. (2022). Affordances and challenges of artificial intelligence in K-12 education: a systematic review. *Journal of Research on Technology in Education*, <https://doi.org/10.1080/15391523.2022.2121344>
- Dabisch, V. (2023) The practices of data-based governance: German school supervision, professionalism and datafied structurations. *Tertium Comparationis*, 29(1), 48-72.
- Davies, H., Eynon, R., Komljenovic, J. and Williamson, B. (2022) Investigating the financial power brokers behind EdTech. In Livingstone, S. and Pothong, K. (eds). *Education Data Futures: Critical, Regulatory and Practical Reflections*. 5Rights Foundation.
- Day, E., Pothong, K., Atabey, A. and Livingstone, S. (2022) Who controls children's education data? A socio-legal analysis of the UK governance regimes for schools and EdTech. *Learning, Media and Technology* <https://doi.org/10.1080/17439884.2022.2152838>
- Decuyper, M., Grimaldi, E. and Landri, P. (2021) Introduction: Critical studies of digital education platforms. *Critical Studies in Education* 62(1): 1-16.
- Digital Futures Commission. 2023. *A Blueprint for Education Data: Realising children's best interests in digitised education*. Digital Futures Commission, March 2023: <https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/A-Blueprint-for-Education-Data-FINAL-Online.pdf>
- Ebbeler, J., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2016). Effects of a data use intervention on educators' use of knowledge and skills. *Studies in educational evaluation*, 48, 19-31.
- Eynon, R. 2022. Datafication and the role of schooling: Challenging the status quo. In

- Pangrazio, L. and Sefton-Green, J. (eds). *Learning to Live with Datafication: Educational case studies and initiatives from across the world*. Routledge.
- Faber, J. M., Luyten, H., & Visscher, A. J. (2017). The effects of a digital formative assessment tool on mathematics achievement and student motivation: Results of a randomized experiment. *Computers & Education*, 106, 83–96
- García-Senín, S., Arguedas, M., & Daradoumis, T. (2022). Using learning analytics to support STEAM students' academic achievement and self-regulated learning. *REM : Research on Education and Media*, 14(1), 36–45. <https://doi.org/10.2478/rem-2022-0005>
- Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *TechTrends*, 59(1), 64–71. <https://doi.org/10.1007/s11528-014-0822-x>
- Grek, S., Maroy, C. and Verger, A. (eds). 2021. *Accountability and Datafication in the Governance of Education*. London: Routledge.
- Gulson, K.N. & Witzemberger, K. 2020. Repackaging authority: Artificial intelligence, automated governance and education trade shows. *Journal of Education Policy*: <https://doi.org/10.1080/02680939.2020.1785552>
- Hakimi, L., Eynon, R. and Murphy, V.A. (2021) The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. *Review of Educational Research*: <https://doi.org/10.3102/00346543211020116>
- Hansen, M. & Komljenovic, J. (2023) Automating Learning Situations in EdTech: Techno Commercial Logic of Assetisation. *Postdigital Science and Education*, 5, 100–116
- Hartong, S. (2018) Towards a topological re-assemblage of education policy? Observing the implementation of performance data infrastructures and 'centers of calculation' in Germany. *Globalisation, Societies and Education*, 16(1), 134–150.
- Hartong, S. and Decuyper, M. (2023) Platformed professional(itie)s and the ongoing digital transformation of education. *Tertium Comparationis*, 29(1), 1-21.
- Hartong, S. and Förschler, A. 2019. Opening the black box of data-based school monitoring: Data infrastructures, flows and practices in state education agencies. *Big Data & Society*, 6(1). <https://doi.org/10.1177/2053951719853311>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112. <https://doi.org/10.3102/003465430298487>
- Helsper, E. J. (2021). *The digital disconnect : the social causes and consequences of digital inequalities*. SAGE.
- Henne, T. & Gstrein, O.J. (2023) forthcoming. Governing the 'datafied' school: Bridging the divergence between universal education and student autonomy. In Zittler, A. and Gstrein, O.J. (eds). *Handbook on the Politics and Governance of Big Data and Artificial Intelligence*. Cheltenham: Edward Elgar.
- Hebbecker, K., Förster, N., Forthmann, B., & Souvignier, E. (2022). Data-based decision-making in schools: Examining the process and effects of teacher support. *Journal of Educational Psychology*, 114(7), 1695–1721. <https://doi.org/10.1037/edu0000530>

- Hintz, A., Dencik, L., & Wahl-Jorgensen, K. (2017). Digital citizenship and surveillance society: Introduction. *International Journal of Communication* 11, 731–739.
- Holmes, W., Persson, H., Chounta, I.-A., Wasson, B., & Dimitrova, V. (2022). Artificial Intelligence and education: A critical view through the lens of human rights, democracy, and the rule of law. Council of Europe.  
<https://rm.coe.int/prems-092922-gbr-2517-ai-and-education-txt-16x24-web/1680a956e3>
- Hoogland, I., Schildkamp, K., van der Kleij, F., Heitink, M., Kippers, W., Veldkamp, B., & Dijkstra, A. M. (2016). Prerequisites for data-based decision making in the classroom: Research evidence and practical illustrations. *Teaching and Teacher Education*, 60, 377–386.  
<https://doi.org/10.1016/j.tate.2016.07.012>
- Hooper, L., Livingstone, S., and Pothong, K. (2022) Problems with data governance in UK schools: the cases of Google Classroom and ClassDojo. *Digital Futures Commission*, 5Rights Foundation.
- Hopkins, D., Stringfield, S., Harris, A., Stoll, L., and Mackay, T. (2014). School and System Improvement: a Narrative State-Of-The-Art Review. *Sch. Effectiveness Sch. Improvement* 25 (2), 257–281. <https://doi.org/10.1080/09243453.2014.885452>
- Howard, S. K., Swist, T., Gasevic, D., Bartimote, K., Knight, S., Gulson, K., Apps, T., Peloché, J., Hutchinson, N., & Selwyn, N. (2022). Educational data journeys: Where are we going, what are we taking and making for AI? *Computers and Education. Artificial Intelligence*, 3, 100073. <https://doi.org/10.1016/j.caeai.2022.100073>
- Hubers, M. D., Poortman, C. L., Schildkamp, K., & Pieters, J. M. (2019). Spreading the word: Boundary crossers building collective capacity for data use. *Teachers college record*, 121(1), pp. 1-45.
- Iliadis, A. & Russo, F. (2016). Critical data studies: An introduction. *Big Data & Society*, 3(2) <https://doi.org/10.1177/2053951716674238>
- Imants, J., & Van der Wal, M. M. (2020). A model of teacher agency in professional development and school reform. *Journal of Curriculum Studies*, 52(1), 1-14.
- Jahnke, I. & A. Norberg (2013). Digital Didactics: Scaffolding a New Normality of Learning. In: *Open Education 2030: contributions to the JRC-IPTS Call for Vision Papers. Part III: Higher Education*. pp. 129-134.
- Jarke, J., & Breiter, A. (2020). *The Datafication of Education*. Taylor & Francis.
- Jasanoff, S. (2015). Future imperfect: science, technology, and the imaginations of modernity. In Jasanoff, S. and Kim, S-H. (eds), *Dreamscapes of Modernity: sociotechnical imaginaries and the fabrication of power*. Chicago, IL: University of Chicago Press, pp. 1-33.
- Jenks, C. (2005). (Ed.). *Childhood*. Routledge.
- Jung, P., McMaster, K. L., Kunkel, A. K., Shin, J., & Stecker, P. M. (2018). Effects of data-based individualization for students with intensive learning needs: A meta-analysis. *Learning Disabilities Research & Practice*, 33, 144–155. <https://doi.org/10.1111/lrdp.12172>
- Kerr, K. A., Marsh, J. A., Ikemoto, G. S., Darilek, H., & Barney, H. (2006). Strategies to

- promote data use for instructional improvement: Actions, outcomes, and lessons from three urban districts. *American Journal of Education*, 112(4), 496–520.  
<https://doi.org/10.1086/505057>
- Kerssens, N. and van Dijck, J. (2022) Governed by Edtech? Valuing Pedagogical Autonomy in a Platform Society. *Harvard Educational Review*, 92(2), 284–303.
- Kerssens, N., & van Dijck, J. (2021). The platformization of primary education in The Netherlands. *Learning, Media and Technology*, 46(3), 250–263.  
<https://doi.org/10.1080/17439884.2021.1876725>
- Kippers, W. B., Wolterinck, C. H. D., Schildkamp, K., Poortman, C. L., & Visscher, A. J. (2018). Teachers' views on the use of assessment for learning and data-based decision making in classroom practice. *Teaching and Teacher Education*, 75(October), 199–213.  
<https://doi.org/10.1016/j.tate.2018.06.015>
- Kizilcec, R. F. & Lee, H. (2022). Algorithmic Fairness in Education. In W. Holmes & K. Porayska-Pomsta (Eds.), *The Ethics of Artificial Intelligence in Education*, Routledge.
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education (Columbia, S.C.)*, 22(3), 255–276.  
<https://doi.org/10.1080/10476210.2011.593164>
- Komljenovic, J. (2021) The rise of education rentiers: digital platforms, digital data and rents, *Learning, Media and Technology*, 46(3), 320–332.
- Kordzadeh, N., & Ghasemaghahi, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409.  
<https://doi.org/10.1080/0960085X.2021.1927212>
- Krein, U., & Schiefner-Rohs, M. (2021). Data in Schools: (Changing) Practices and Blind Spots at a Glance. *Frontiers in Education (Lausanne)*, 6  
<https://doi.org/10.3389/feduc.2021.672666>
- Kurcikova, N. (2022). The promise and pitfalls of personalised learning with new EdTech. In Digital Futures Commission, UK.  
<https://educationdatafutures.digitalfuturescommission.org.uk/>
- Kumar, P.C., Chetty, M., Clegg, T., & Vitak, J. (2019). Privacy and security considerations for digital technology use in elementary schools. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. <https://doi.org/10.1145/3290605.330053>
- Landri, P. (2018) *Digital Governance of Education*. Bloomsbury.
- Lawn, M. 2013. The Rise of Data in Education. In Lawn, M. (ed.). *The Rise of Data in Education Systems: Collection, Visualization and Use*, 7–25. Symposium.
- Leeuwen, A. van, Knoop-van Campen, C. A. ., Molenaar, I., & Rummel, N. (2021). How teacher characteristics relate to how teachers use dashboards: Results from two case studies in K-12. *Journal of Learning Analytics*, 8(2), 6–21.  
<https://doi.org/10.18608/JLA.2021.7325>

- Lieberman, M. (2020). Massive shift to remote learning prompts big data privacy concerns. *EducationWeek*. Accessed from <https://www.edweek.org/technology/massive-shift-to-remote-learning-prompts-big-data-privacy-concerns/2020/03>
- Livingstone, S. (2020). Realizing children’s rights in relation to the digital environment. *European Review*. Retrieved from <http://eprints.lse.ac.uk/103563/>
- Livingstone, S., Stoilova, M. & Nandagiri, R. (2020) Data and privacy: the role of the school in educating children in a datafied society. In: Frau-Meigs, Divina, Kotilainen, Sirku, Pathak-Shelat, Manisha, Hoechsmann, Michael and Poyntz, Stuart R., (eds.) *The Handbook of Media Education Research. Global Handbooks in Media and Communication Research*. Wiley-Blackwell, Hoboken, New Jersey, 413 - 425. ISBN 9781119166870
- Livingstone, S. & Pothong, K. (2022). The problem and the potential of children's education data. In *Digital Futures Commission Education Data Futures. Critical, Regulatory and Practical Reflections*. <https://educationdatafutures.digitalfuturescommission.org.uk/>
- Lockton, M., Weddle, H. & Datnow, A. (2020) When data don’t drive: teacher agency in data use efforts in low-performing schools, *School Effectiveness and School Improvement*, 31:2, 243-265, <https://doi.org/10.1080/09243453.2019.1647442>
- Lomborg, S., & Ytre-Arne, B. (2021). Advancing digital disconnection research: Introduction to the special issue. *Convergence*, 27(6), 1529-1535. <https://doi.org/10.1177/13548565211057518>
- Luo, J., Wang, M., & Yu, S. (2022). Exploring the factors influencing teachers’ instructional data use with electronic data systems. *Computers & Education*, 191, 104631.
- 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>
- Malinka, K., Perešini, M., Firc, A., Hujňák, O., & Januš, F. (2023). On the Educational Impact of ChatGPT: Is Artificial Intelligence Ready to Obtain a University Degree? <https://doi.org/10.48550/arxiv.2303.11146>
- Mandinach, E. B., & Gummer, E. S. (2016). What does it mean for teachers to be data literate: Laying out the skills, knowledge, and dispositions. *Teaching and Teacher Education*, 60, 366–376. <https://doi.org/10.1016/j.tate.2016.07.011>
- Mascheroni, G. (2018). Datafied childhoods: Contextualising datafication in everyday life. *Current Sociology*, 1-16. <https://doi.org/10.1177/0011392118807534>
- Mikaite, I., De Wolf, R., Lievens, S., De Leyn, T., & Marten, M. (2021). Children’s reflections on privacy and the protection of their personal data: A child-centric approach to data protection information formats. *Children and Youth Services Review*, 129, 106170. <https://doi.org/10.1016/j.childyouth.2021.106170>
- Milkaite, I., & Lievens, E. (2020). Child-friendly transparency of data processing in the EU: from legal requirements to platform policies. *Journal of Children and Media*, 14(1), 5–21. <https://doi.org/10.1080/17482798.2019.1701055>
- Mills, M., Mockler, N., Stacey, M., & Taylor, B. (2021). Teachers' orientations to educational

- research and data in England and Australia: implications for teacher professionalism. *Teaching Education (Columbia, S.C.)*, 32(1), 77–98. <https://doi.org/10.1080/10476210.2020.1843617>
- Molenaar, I., Horvers, A., Dijkstra, R. & Baker, R.S. (2020). Personalized visualizations to promote young learners' SRL: the learning path app. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20)*. Association for Computing Machinery, New York, NY, USA, 330–339. <https://doi.org/10.1145/3375462.3375465>
- Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, 57(4), 632–645. <https://doi.org/10.1111/ejed.12527>
- [Nazaretsky](#), T., [Ariely](#), M., [Cukurova](#), M. & [Alexandron](#), G. (2022). Teachers' trust in AI-powered educational technology and a professional development program to improve it. *British Journal of Educational Technology*, vol. 53, Issue 4, pp. 914-931.
- Nichols, T.P. and Garcia, A. (2022) Platform studies in education. *Harvard Educational Review*, 92(2), 209-230.
- Nichols, T. P., & Dixon-Román, E. (2024). Platform Governance and Education Policy: Power and Politics in Emerging Edtech Ecologies. *Educational Evaluation and Policy Analysis*, 46(2), 309-328. <https://doi.org/10.3102/01623737231202469>
- Niemi, H., Pea, H., Lu, R. D., & Lu, Yu. (2022). *AI in Learning: Designing the Future*. Springer Nature.
- Nordic Innovation. (2022). The Nordic AI and data ecosystem 2022.
- Nottingham, E., Stockman, C., & Burke, M. (2022). Education in a datafied world: Balancing children's rights and school's responsibilities in the age of Covid 19. *Computer Law & Security Review*, 45, 105664.
- OECD (2023). *OECD DIGITAL EDUCATION OUTLOOK 2023. Towards an Effective Digital Education Ecosystem*. Paris: OECD.
- Oslund, E. L., Elleman, A. M., & Wallace, K. (2021). Factors related to data-based decision-making: Examining experience, professional development, and the mediating effect of confidence on teacher graph literacy. *Journal of Learning Disabilities*, 54(4), 243-255.
- Ozga, J., Dahler-Larsen, P., Segerholm, C. and Simola, H. (eds). (2011) *Fabricating Quality in Education: Data and Governance in Europe*. Routledge.
- Pangrazio, L., & Sefton-Green, J. (2022). *Learning to live with datafication : educational case studies and initiatives from across the world*. Routledge.
- Pangrazio, L., Selwyn, N. and Cumbo, B. (2023) A patchwork of platforms: mapping data infrastructures in schools. *Learning, Media and Technology*, 48(1), 65-80.
- Perrotta, C., Gulson, K. N., Williamson, B., & Witzemberger, K. (2020). Automation, APIs and the distributed labour of platform pedagogies in Google Classroom. *Critical Studies in Education*, 62(1), 97–113. <https://doi.org/10.1080/17508487.2020.1855597>
- Perrotta, C., & Selwyn, N. (2020). Deep learning goes to school: toward a relational

- understanding of AI in education. *Learning, Media and Technology*, 45(3), 251–269.  
<https://doi.org/10.1080/17439884.2020.1686017>
- Persson, J. (2022) Ten-years after National Pupil Database commercialisation: ICO reprimands the Department for Education over “woeful” misuse. Defend Digital Me, 6 November: <https://defenddigitalme.org/2022/11/06/ten-years-after-national-pupil-database-commercialisation-ico-reprimands-the-department-for-education-over-woeful-misuse/>
- Piattoeva, N. & Boden, R. 2020. Escaping numbers? The ambiguities of the governance of education through data. *International Studies in Sociology of Education*, 29 (1–2): 1–18.
- Poell, T., Nieborg, D., & van Dijck, J. (2019). Platformisation. *Internet Policy Review*, 8(4).  
<https://doi.org/10.14763/2019.4.1425>
- Priestley, M., Edwards, R., Priestley, A., & Miller K. (2012). Teacher Agency in Curriculum Making: Agents of Change and Spaces for Manoeuvre. *Curriculum Inquiry*, 42(2), 191–214.  
<https://doi.org/10.1111/j.1467-873X.2012.00588.x>
- Prinsloo, P. (2018). Include us all! Directions for the adoption of learning analytics in the GlobalSouth: An African perspective. Digital Learning for Development (DL4D). Available at: <http://dl4d.org/portfolio-items/learning-analytics-for-the-global-south/>
- Ramaswami, G., Susnjak, T., Mathrani, A., & Umer, R. (2022). Use of Predictive Analytics within Learning Analytics Dashboards: A Review of Case Studies. *Technology, Knowledge and Learning*. <https://doi.org/10.1007/s10758-022-09613-x>
- Rauf, D. (2020). Artificial intelligence in K-12 education: Unintended consequences lurk, report warns. Retrieved from <https://www.edweek.org/technology/artificial-intelligence-in-k-12-education-unintended-consequences-lurk-report-warns/2020/05>
- Reeves, T. D., & Honig, S. L. (2015). A classroom data literacy intervention for pre-service teachers. *Teaching and Teacher Education*, 50, 90–101.
- Regan, P. M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking. *Ethics and Information Technology*, 21(3), 167–179. <https://doi.org/10.1007/s10676-018-9492-2>
- Reynolds, D., Sammons, P., De Fraine, B., van Damme, J., Townsend, T., Teddlie, C., et al. (2014). Educational Effectiveness Research (EER): a State-Of-The-Art Review. *Sch. Effectiveness Sch. Improvement* 25 (2), 197–230.  
<https://doi.org/10.1080/09243453.2014.885450>
- Robinson, S. (2020). Trust, transparency, and openness: How inclusion of cultural values shapes Nordic national public policy strategies for artificial intelligence (AI). *Technology in Society*, 63. <https://doi.org/10.1016/j.techsoc.2020.101421>
- Salas-Pilco, S. Z., Xiao, K., & Hu, X. (2022). Artificial Intelligence and Learning Analytics in Teacher Education: A Systematic Review. *Education Sciences*, 12(8), 569.  
<https://doi.org/10.3390/educsci12080569>
- Srnicek, N. (2017). *Platform capitalism*. Polity.
- Sefton-Green, J. & Pangrazio, L. (2022) The death of the educative subject? The limits of

- criticality under datafication, *Educational Philosophy and Theory*, 54:12, 2072-2081, <https://doi.org/10.1080/00131857.2021.1978072>
- Selwyn, N. (2020). Re-imagining 'Learning Analytics' ... a case for starting again? The Internet and Higher Education, 46, 100745. <https://doi.org/10.1016/j.iheduc.2020.100745>
- Selwyn, N., Hillman, T., Bergviken-Rensfeldt, A. & Perrotta, C. (2023) Making Sense of the Digital Automation of Education. *Postdigital Science and Education*, 5, 1–14.
- Schildkamp, K. & Datnow, A. (2022) When Data Teams Struggle: Learning from Less Successful Data Use Efforts, *Leadership and Policy in Schools*, 21:2, 147-166, <https://doi.org/10.1080/15700763.2020.1734630>
- Schildkamp, K., & Kuiper, W. (2010). Data-informed curriculum reform: Which data, what purposes, and promoting and hindering factors. *Teaching and Teacher Education*, 26(3), 482–496. <https://doi.org/10.1016/j.tate.2009.06.007>
- Schildkamp, K., Lai, M. K., & Earl, L. (2013). *Data-based Decision Making in Education: Challenges and Opportunities* (1st ed. 2013., Vol. 17). Springer Netherlands : Imprint: Springer.
- Schildkamp, K., & Poortman, C. (2015). Factors influencing the functioning of data teams. *Teachers College Record*, 117(4), 1–42.
- Schildkamp, K., Poortman, C. L., & Handelzalts, A. (2016). Data teams for school improvement. *School effectiveness and school improvement*, 27(2), 228-254.
- Schildkamp, K., Poortman, C. L., Ebbeler, J., and Pieters, J. M. (2019). How School Leaders Can Build Effective Data Teams: Five Building Blocks for a New Wave of Data-Informed Decision Making. *J. Educ. Change* 20, 283–325. <https://doi.org/10.1007/s10833-019-09345-3>
- Siemens, G., & Long, P. (2011). Penetrating the Fog: Analytics in Learning and Education. *EDUCAUSE Review*, 46(5), 30.
- Siibak, A., & Mascheroni, G. (2021). Children's data and privacy in the digital age. SSOAR – GESIS Leibniz Institute for the Social Sciences. <https://doi.org/10.21241/ssoar.76251>
- Stoilova, M., Livingstone, S., & Nandagiri, R. (2020). Digital by default: Children's capacity to understand and manage online data and privacy. *Media and Communication*, 8, 4, 197-207
- Stoilova, M., Nandagiri, R., & Livingstone, S. (2021). Children's understanding of personal data and privacy online - a systematic evidence mapping. *Information, Communication & Society*, 24(4), 557–575. <https://doi.org/10.1080/1369118X.2019.1657164>
- Stornaiuolo, A., Higgs, J., Nichols, T.P., Leblanc, R.J. & de Roock, R.S. (2023). The Platformization of Writing Instruction: Considering educational equity in new learning ecologies. *Review of Research in Education*. Vol. 47, pp. 311-359
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167. <https://doi.org/10.1016/j.chb.2014.05.038>
- Tondeur, J., Sinnaeve, I., van Houtte, M., & van Braak, J. (2011). ICT as cultural capital: The

- relationship between socioeconomic status and the computer-use profile of young people. *New media & society*, 13(1), 151-168.
- UNESCO (2018). General principles on children's online privacy and freedom of expression. Industrytoolkit. [https://sites.unicef.org/csr/files/UNICEF\\_Childrens\\_Online\\_Privacy\\_and\\_Freedom\\_of\\_Expression\(1\).pdf](https://sites.unicef.org/csr/files/UNICEF_Childrens_Online_Privacy_and_Freedom_of_Expression(1).pdf)
- UNESCO (2022). *Minding the data: protecting learner's privacy and security*. <https://unesdoc.unesco.org/ark:/48223/pf0000381494>
- 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>
- van Dijck, J. (2020). Governing digital societies: Private platforms, public values. *The Computer Law and Security Report*, 36, 105377. <https://doi.org/10.1016/j.clsr.2019.105377>
- van Dijck, J. (2021) Seeing the forest for the trees: Visualizing platformization and its governance. *New Media & Society*, 23(9), 2801–2819.
- van Dijck, J., Poell, T and de Waal, M. (2018) *The Platform Society: Public values in a connective world*. Oxford University Press.
- van Krieken R. (2018) *Celebrity Society: The Struggle for Attention*, 2nd edn. London: Routledge.
- Vanlommel, K., van Gasse, R., Vanhoof, J., & van Petegem, P. (2017). Teachers' decision-making: Data based or intuition driven? *International Journal of Educational Research*, 83, 75–83. <https://doi.org/10.1016/j.ijer.2017.02.013>
- Vartiainen, H., Toivonen, T., Jormanainen, I., Kahila, J., Tedre, M., & Valtonen, T. (2021). Machine learning for middle schoolers: Learning through data-driven design. *International Journal of Child-Computer Interaction*, 29, 1–12. <https://doi.org/10.1016/j.ijcci.2021.100281>
- Veale, M. (2022) Schools must resist EdTech, but it won't be easy. In Livingstone, S. and Pothong, K. (eds). *Education Data Futures: Critical, Regulatory and Practical Reflections*. 5Rights Foundation.
- Verger, A., Lubienski, C., & Steiner-Khamsi, G. (eds.) (2016) *World yearbook of education 2016: The global education industry*. Routledge.
- Verger, A., Parcerisa, L. and Fontdevila, C. (2019) The growth and spread of large-scale assessments and test-based accountabilities: a political sociology of global education reforms. *Educational Review*, 71(1), 5-30.
- Vezzoli, Y., Mavrikis, M., & Vasalou, A. (2020, March). Inspiration cards workshops with primary teachers in the early co-design stages of learning analytics. In Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (pp. 73-82). <https://doi.org/10.1145/3375462.3375537>
- Viberg, O. & Grönlund, A. (2023). Introducing Practicable Learning Analytics. <https://doi.org/10.48550/arxiv.2301.13043>
- Viberg, O., Kizilcec, R., Wise, A., Jivet, I., & Nixon, N. (2024). Advancing equity and inclusion

- in educational practices with AI-powered educational decision support system (AI-EDSS). *British Journal of Educational Tecnology*, 55(5), 1974-1981. <https://doi.org/10.1111/bjet.13507>
- Wall, K., Higgins, S., Remedios, R., Rafferty, V., & Tiplady, L. (2013). Comparing Analysis Frames for Visual Data Sets. *Journal of Mixed Methods Research*, 7(1), 22–42. <https://doi.org/10.1177/1558689812450211>
- Wasson, B. & Kirschner, P. (2020). *Learning Design: European Approaches*. <https://doi.org/https://doi.org/10.1007/s11528-020-00498-0>
- Wasson, B. Hansen, C. & Netteland, G. (2016). *Data literacy and use for learning when using learning analytics for learners*. <http://hdl.handle.net/11250/2429460>
- Wayman, J. C., Midgley, S., & Stringfield, S. (2017). Leadership for data-based decision making: Collaborative educator teams. In *Learner-centered leadership* (pp. 189-206). Routledge.
- Williamson, B. (2016a) Digital education governance: data visualization, predictive analytics and 'real-time' policy instruments. *Journal of Education Policy*, 31(2), 123-141.
- Williamson, B. (2016b) Digital education governance: an introduction. *European Educational Research Journal*, 15(1), 3-13.
- Williamson, B. (2016c) Digital methodologies of education governance: Pearson plc and the remediation of methods. *European Educational Research Journal*, 15(1): 34-53.
- Williamson, B. (2017). *Big data in education : the digital future of learning, policy and practice*. SAGE
- Williamson, B. (2019). Policy networks, performance metrics and platform markets: Charting the expanding data infrastructure of higher education. *British Journal of Educational Technology*, 50(6), 2794–2809. <https://doi.org/10.1111/bjet.12849>
- Williamson, B. (2023) Governing through infrastructural control: artificial intelligence and cloud computing in the data-intensive state. In Housley, W., Edwards, A., Beneito-Montagut, R. & Fitzgerald, R. (eds). *The Sage Handbook of Digital Society*. Sage
- Williamson, B. & Piattoeva, N. (2019) Objectivity as standardization in data scientific educational policy and governance. *Learning, Media & Technology*, 44(1), 64-76.
- Williamson, B., Eynon, R., & Potter, J. (2020). Pandemic politics, pedagogies and practices: digital technologies and distance education during the coronavirus emergency. *Learning, Media and Technology*, 45(2), 107–114. <https://doi.org/10.1080/17439884.2020.1761641>
- Williamson, B., Gulson, G., Perrotta, C. & Witzemberger, K. (2022) Amazon and the new global connective architectures of educational governance. *Harvard Educational Review*, 92(2), 231–256.
- Williamson, B., Komljenovic, J., & Gulson, K. N. (2024). Introduction: Digitalisation of education in the era of algorithms, automation and artificial intelligence. In B. Williamson, J. Komljenovic, & K. N. Gulson (Eds.), *World Yearbook of Education 2024: Digitalisation of Education in the Era of Algorithms, Automation and Artificial Intelligence* (pp. 1–19). London: Routledge.
- Witzemberger, K. & Gulson, K.N. (2021). Why EdTech is always right: students, data and

- machines in pre-emptive configurations. *Learning, Media and Technology*, 46(4), 420-434.
- Yan, L., Martinez-Maldonado, R., & Gasevic, D. (2024). Artificial intelligence in learning analytics: Contextualising opportunities and challenges through the learning analytics cycle. In *Proceedings of the 14th Learning Analytics and Knowledge Conference*, pp. 101-111. <https://doi.org/10.1111/bjet.13507>
- Yu, J., & Couldry, N. (2022). Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. *Education, Communication & Society*, 25(1), 127-144.
- Zuboff, S. (2015). Big other: Surveillance Capitalism and the Prospects of an Information Civilization. *Journal of Information Technology*, 30(1), 75–89. <https://doi.org/10.1057/jit.2015.5>
- Zuboff, S. (2019) *The age of surveillance capitalism: The fight for a human future at the new frontier of power: Barack Obama's books of 2019*. Profile books.



**Co-funded by  
the European Union**

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.