Meaningful and ethical use of data in schools
The front cover was inspired by the defragmentation tool, used to organise files within hard drives. It is also made to look like a server, where data is stored.

Special thanks goes to Katja Engelhardt who moderated and summarised the discussions of the first webinar that took place in October 2022.
CONTENTS

01  Introduction

04  (How) is using data for learning ethical?

11  The platformisation of education, capturing data & testing

14  What does the use of algorithms mean for the autonomy of teachers?

18  How can teachers best use and make sense of data in education?

23  ChatGPT and the role of AI in assessment

30  Governance processes for a robust educational data ecosystem
Introduction

Data4Learning: key issues for further investigation and development.

In the ever-evolving discourse on the purpose and role of education, the debate about the use of digitally processed data has been growing in the past years. This debate has drawn particular attention in the post-COVID period that emerged after the reopening of schools across the world and the intensification of technology use across most educational systems. In an effort to contribute to this discussion, European Schoolnet launched in late 2022 a series of webinars, titled Data4Learning, inviting key experts and professionals from education authorities and academic institutions to exchange, reflect, and share their insights over six episodes from October 2022 to June 2023. These episodes delve into the complexities and potentials of the increasing production of data in education and cover topics about the platformisation of education, the use of algorithms and the impact on teachers’ autonomy, the teachers’ sensemaking of data use in education, generative Artificial Intelligence (AI) and assessment, ethics, and governance of data ecosystems. The current Introductory text takes stock of the key messages shared during the webinars and outlines identified key issues for further investigation and development. The six webinar summaries have been individually published over the course of the series’ duration but are also presented in this report chronologically for the readers’ convenience and their further consideration.

Datafication in education

One of the main themes that emerged during the webinar series refers to the phenomenon of datafication in education. The use of algorithmic technology in schools increasingly facilitates data-driven practices which involve using various data sources to gain insights into students’ learning and teaching methods. The large amount of data collected in educational settings is often argued to be the most reliable way for schools to have a comprehensive view of things. The four types of data collection that are most prevalent in education settings include small-scale teacher-collected data (also called ‘micro-data’), large external sources like national tests, learning analytics data, and trace data collected automatically through the use of digital technology such as button clicks or response times during online assignments. Nevertheless, teachers also acquire insights into their students’ learning through classroom interactions, observations, and multiple other types of situations. The use of data can complement this process by raising awareness of the way students engage in the learning process. Such practices also offer the potential to tailor educational approaches to individual or group needs.

However, some of the limitations include the reductiveness and oversimplification of what is measured, and the misrepresentation of some aspects that data cannot capture. Moreover, this tendency raises questions around the impact that datafication in education can have on students’ learning and the purpose of education overall. Furthermore, it is argued that the -often-observed- use of multiple platforms which collect student data can complicate teachers’ understanding and ability to effectively use their insights. The use of data can sometimes also create an imbalanced dynamic, where students and teachers provide information without fully understanding its generation or purpose. This relates to the term “surveillance pedagogy” which highlights concerns about tracking and monitoring students through technology, limiting individual agency and data protection. The level of automation and the teacher-algorithm relationship is argued to be a delicate one, as teachers are most often those accountable for decisions in their classroom. Besides, digital platforms that collect data can often identify problems, but they do not always provide solutions for improvement or adaptation in a given situation. Developing algorithmic literacy and critical AI literacy is considered important in such datafied settings while it is also crucial that the goal of using such systems remains the will to harness AI systems to augment teaching in innovative ways beyond mere replication of teachers’ work.
Data in education and assessment

Another important theme that came up several times during the discussions with the invited experts relates to student assessment. The shift towards online evaluations and adaptive learning platforms is argued to open a plethora of opportunities in education, but also to raise significant concerns that need to be addressed. On the one hand, online platforms used in education often promise to allow students the flexibility to learn at their own pace, while also allowing teachers benefit from automated tasks like grading, enabling them to focus on instructional planning and individualised support. Such opportunities, however, can only be harnessed if teachers are able to develop new skills to effectively interpret students’ information that is provided through these platforms. Enhancing teachers’ data literacy can foster their understanding on the value and possibilities that these tools offer. It has been noticed that experienced teachers rely more on their judgment, having been in the classroom for longer, while less experienced teachers can often rely more on the use of data to validate their knowledge.

This need to address the concerns raised is even more pressing at a time when the use generative AI tools threaten the traditional methods of assessment that many schools still employ to evaluate their students. Although some have advocated for a total ban on generative AI tools in schools, others argue that such move would not solve the current challenges and could even widen inequalities among students. When it comes to solutions, there are short-, medium-, and longer-term possibilities proposed. In the short-term, assessment changes could involve oral or invigilated exams in schools and focus on teacher professional development in the area of AI literacy. Mid-term changes could emphasise project-based assessment which can limit the value of using AI to produce specific work. The more long-term possibilities include the introduction of formative assessment methods which can empower students’ learning. Use of AI-tools is argued to support the personalisation of education, ensuring equal support to all students, provided that the tools are reliable. AI is also argued to free-up teachers’ time and complement their work. However, AI cannot be seen as a possible replacement for teachers’ important role in education.

Schools, teachers, and data ecosystems

Data ecosystems in education operate on multiple levels, involving various actors and structures on micro, meso, and macro levels. These ecosystems include various actors, such as students, teachers, parents, technology providers, and various education authorities, alongside structures such as current laws and regulations. A whole-school approach where stakeholders engage in conversations about the use of technology in the school, including during procurement processes, could possibly address many of the risks and questions at hand. However, one should not ignore that technology providers and schools often have different objective with the former having the technical control over the data processed and the latter carrying the responsibility to ensure that the tools they use only process data strictly according to their own defined purposes. Schools often feel pressured to integrate technology, but the large number of options and the lack of strong evidence of their value can lead to confusion and inaction. In that struggle, education authorities have arguably a role to play in creating the right structures to evaluate and filter the technology integrated in schools and to ensure interoperability of data collected across tools and schools which can improve evaluation of indicators such as performance, quality, and accountability.

Moreover, education authorities should make sure that the right checks and balances are introduced already at design stage to safeguard data accuracy, children rights, and diversity of training data.

In their capacity, schools entering into agreements about the collection and processing of their students’ information, need to consider both types of data that are visible to the school and hidden data that might be collected in the background for commercial purposes. Hidden data raises significant transparency issues since schools are often not aware what student data is being collected and how it is used. The question of ownership of data belonging to minors further complicates the role and responsibility of the schools considering the premise that “ownership entails control”.

Furthermore, the discussion about student data ownership involves multiple stakeholders, including the students, their parents, but also the schools and educational authorities which access and process their data. Empowering schools and teachers to properly understand and use data, and clearly defining data collection purposes can help them better interpret and use data to make informed decisions.
Ethics of data and artificial intelligence

Invited speakers denoted ‘ethical’ in education as something that safeguards human rights for learners, teachers, and stakeholders. Ethical technology comprises legal frameworks, value-based education approaches, and ethical responsibility. However, an important linguistic distinction would be the use of “ethics of data” or “data ethics” rather than “ethical data” or “ethical AI” which avoids attributing agency to data or AI systems. Data issues emerge throughout the processes of collection, storage, analysis, and interpretation. On the other hand, schools often do not fully understand what data is collected and how it is processed through the tools used in classrooms. To exercise more control over the data collected, schools ought to receive proper support and incentives to take up the responsibility and develop the expertise needed.

When it comes to the role of teachers, there is a question of reducing, or reconsidering, accountability targets for educators which can enhance meaningful engagement with data. Webinar speakers agree that educators require pragmatic criteria when selecting technology, and educational technology companies need to prioritise educators’ needs. However, there is a delicate balance between the involvement of the public and private sectors in education. Commercial platforms often collect student data without proper consent which restricts schools and students control of the type and volume of data that is collected. This control is further limited when schools do not offer alternatives to students, making the possibility to opt-out, and therefore control, not a real choice.

Moreover, the use of adaptive learning platforms has been criticised to contribute to the increasing phenomena of commercialisation and privatisation of education, especially when there is limited independent evidence at scale that show these technology tools have any positive impact. Platform owners often rely on a limited number of cloud infrastructure providers that have eventually access to the data collected, a process which further obscures the transparency of the systems and the question of data ownership.

Such concerns are exacerbated with the growing use of generative AI models, often freely available online and used by students. Generative AI tools are based on neural networks trained on vast amounts of textual content from the internet, producing text, images, and videos to create novel content. Generative AI systems, like ChatGPT, leverage existing media to produce new content. These models often rely on outdated training data which mirror societal issues and raise concerns about biases, racism, and equal access for all users. Inherent biases in the data and questions of system ownership pose significant ethical dilemmas while the effect of “hallucination” that all generative AI systems fall into can lead to the creation of fabricated information. Education actors and institutions have a duty to care and utilise data to benefit children’s education. Therefore, the use of data in schools should aim to ultimately support student learning.

Conclusion

As we navigate the transformation that the increasing use of technology and AI in education contributes to, it is important to acknowledge the opportunities that the use of these tools and of the magnitude of data produced through their use, but also to understand the challenges and limitations they have, especially when it comes to the unmeasurable aspects in education provision. The themes that emerged from the webinars exactly underscore the importance of embracing opportunities while addressing challenges. Educators, policymakers, and other stakeholders need to further collaborate to harness the potential of data and promote more student-tailored learning, equitable practices, and informed decision-making in education. By prioritising ethics, transparency, and data literacy, the education community can build a future where data and AI serve as tools to enhance learning experiences and empower both students and teachers, rather than limit or replace them. European Schoolnet contributes to this effort as part of its core mission to support ministries of education, schools, teachers, and relevant education stakeholders in Europe in the transformation of education processes for 21st century digitalised societies. It is doing that, more tangibly, through initiatives and activities such as the Data4Learning webinar series and the recently launched AgileEDU project, which is funded by the European Commission and brings together diverse actors from across Europe to investigate the pedagogical use of digitally processes data in education.

European Schoolnet...
The platformisation of education, capturing data & testing

**What does ‘platformisation of education’ mean?**

The term ‘platform’ can refer to online learning spaces in general or to platforms that include adaptive elements. Shifting evaluations online generally means a shift towards more privatisation and commercialisation of education, as Niels Kerssens highlighted. He observed that very similar developments took place in Ireland, Norway, and the Netherlands. While initiatives for publicly funded and/or owned education platforms exist, for example, in the Netherlands and Greece, it is generally difficult to compete with private companies, as substantial and continuous investment is required.

**What are adaptive learning platforms, and what is their potential?**

Marthe Straatemeier explained that the term ‘adaptive learning platform’ usually refers to platforms that adapt to the learning levels of students. Such platforms measure students’ performance continuously and adapt the difficulty of the following tasks based on the student’s responses. In that way they potentially allow each pupil to continue learning at their own pace, which can be very motivating and enable more efficient learning. Teachers also receive a good overview of their students’ learning trajectories, according to Marthe Straatemeier. Niels Kerssens added that companies often claim that teachers gain more autonomy with the use of such platforms, as the automation of tasks like grading students’ work free up teachers’ time.

Different forms of adaptive learning platforms exist, some are offered in addition to usual teaching methods while others provide comprehensive teaching programs covering the entire subject curriculum. More than half of the schools in the Netherlands is estimated to already use platforms that are employing some form of adaptiveness. Adaptive learning platforms started to be used in the Netherlands around ten years ago and Niels Kerssens explained that one can observe a trend of more and more schools using such comprehensive offers.

These learning platforms are very powerful and attractive to schools in Ireland as well, as Tom Lonergan, PDST Technology in Education, Ireland highlighted. They are now one of the most important tools that schools see to be able to support teaching, learning and assessment.
He added that the big data such platforms use raise concerns and questions about access and ownership of information.

What are challenges related to the use of such platforms?

“Coping with adaptive learning platforms is a big challenge for governments and national research.”

Morten Søby

There is a growing concern that the platformisation of education is very powerful and follows its own agenda, as Morten Søby, senior policy advisor, Norwegian Directorate for Education and Training, Norway expressed. He highlighted that while adaptive learning platforms gathered a lot of data, it was not easy for school owners, researchers, or policy makers at national level to get access to that data.

Niels Kerssens stressed that interpreting the information about student learning requires the development of new expertise for teachers and schools. He also emphasised that platform providers need to be more transparent about the back end of their systems, their pedagogical visions, and how their platforms are built, often on bigger cloud infrastructure provided by Google, Amazon, or Microsoft. Marthe Straatemeier agreed that transparency is needed from these companies about their algorithms and their quality.

What are ways forward?

Niels Kerssens put forward that the public sector needs to gain more control over the digitalisation of education in public schools. He stressed that this is a role for education ministers at national level and the European Commission to find ways to regulate and control platforms of big technology companies and even carve out a public space for developing alternative infrastructures for education institutions to store their data. Niels Kerssens suggested that general data privacy agreements that already exist in the Netherlands could be extended to include requirements for more transparency. Marthe Straatemeier advocated for stimulating developers to become more transparent and to engage in a dialogue with researchers and policy makers on how that could work in practice.

The University of Utrecht and Kennisnet collaborate, together with others, on research to provide schools with more insights on how platforms work to enable them to take better informed decisions concerning procurement as well as the use of learning platforms, which are in line with their pedagogical visions. As Marthe Straatemeier explained, Kennisnet supports schools in choosing teaching methods by helping them ask the right questions and develop their own vision. In the Netherlands, a new national artificial intelligence lab just started its’ work which will also look at the use of adaptive technologies in education.

Reflective questions for policy makers

“It is very important to reflect on what the public sector should do.”

Marthe Straatemeier

1. Schools in the Netherlands are very positive about the use of such platforms. Niels Kerssens suggested that policy makers and others need to ask the right questions, e.g.: What does the use of such platforms mean for the autonomy of teachers and schools?

2. Adaptive learning platforms have great potential to enhance learning. Marthe Straatemeier advocated for asking the question: Which type of learning do these technologies really benefit, and which ones do they not?

3. Adaptive learning platforms measure students’ performance continuously. Marthe Straatemeier asked the questions: What do we know about the quality of the measurements that adaptive learning systems use? How much more transparency is needed?

4. Shifting evaluations online will almost always mean a shift towards more commercialisation of education. Niels Kerssens suggested to reflect on the questions: To what extent do education systems want to outsource technological developments to private companies, and what technical expertise is needed to intervene in a meaningful way when it comes to tech companies?

5. In Norway, it is usually school owners who chose the platforms for their schools. Morten Saby, suggested to reflect on the question: What can teachers still decide?
text


What does the use of algorithms mean for the autonomy of teachers?

Data driven practices, as Inge Molenaar explained briefly, can be referred to as “the use of data that helps provide more insights into how students learn and how teachers teach”. This use allows to get better insights into the learning and teaching processes and to improve teaching and learning in an informed way.

Neil Selwyn raised the issue of datafication that comes into play when a lot of different data points are gathered in a school or education setting, exacerbated with the use of technology today. He connected it to what Dave Beer (2019) coined to as the “data gaze”, the increasing tendency for people to rely on data and algorithms to provide a comprehensive view of the world. According to him, there are four types of data in education. The first is small scale data that is context-bound and collected by teachers in their own classrooms. The second refers to data from large external sources (e.g., national tests, PISA, etc.). The third relates to learning analytics data, generated by computers to provide information about students (e.g., dashboards) and the fourth is trace data generated automatically using digital technology.

However, Neil Selwyn mentioned that what teachers consider as data in education is very limited. They seem to be interested mostly in the first type of data. Teachers, according to him, are interested in data that can directly inform them on how to improve their own practice or what to do with the students they currently have in their classroom, and not so much on broader trends or what might happen in 10 years from now. Teachers want to be able to trust the data they are using and as such they can be sceptical of data that are not collected or generated by them. Inge Molenaar also agreed with Neil that teachers usually want to understand what’s going on in their own context and are interested in raw rather than classified data.

What do we mean by data-driven practices in education?

“Teachers are interested in very small-scale data that they generate within their own classrooms to know their own pupils with whom they already have a relationship with; this data is very context bounded.” Neil Selwyn

---

2 “datafication” refers to the increasing use of data and technology in the field of education, particularly in the form of data collection and analysis, to inform and guide decisions about teaching and learning. It involves the generation and use of large amounts of data about students, teachers, and schools to track performance, support planning and decision making, and monitor progress. The concept is linked to broader trends in society towards the “datafication” of many aspects of life, as digital technologies become increasingly prevalent.

3 “classified data” typically refers to sensitive and confidential information about students, teachers, and schools that is collected, stored, and used by educational institutions and organisations. This information can include student demographic information, academic performance data, behavioural and disciplinary records, and other sensitive information. For a more in-depth analysis, you can refer to Molenaar & Knoop van Kampen, 2018 https://ieeexplore.ieee.org/abstract/document/8400475
What are the potential benefits, risks, and limitations that data-driven practices can bring?

According to Inge Molenaar the benefits of data-driven practices in education are the opportunities offered to adjust the educational processes more to the individual needs of students or to the needs of groups of students, to provide educational services that better fit with what students need, to drive their motivation, and to think about some of the larger problems in the educational arena. Tapani Saarinen mentioned that as a teacher, during the COVID19 distance learning period, he had the opportunity to use a digital platform that provided him with learning analytics data. This helped him know what his students were doing and, in his opinion, helped the students to get immediate feedback and support during their tasks. Therefore, a benefit of such practices related to the multitude of information a teacher can have about their students. Tapani Saarinen also mentioned that data-driven practices can improve transparency and accountability in decision making, as well as help identify early learning challenges, but also talent and other individual qualities that students might have.

Neil Selwyn recognises two limitations in data-driven practices: reductiveness and representativeness. The first questions what is over-simplified or essentialised when we measure something by collecting data, while the second refers to what is mis-represented, missed out or simply cannot be represented with data. Neil Selwyn believes that the full range of what happens during the lesson cannot be turned fully into data, because this way we would risk missing “what is so exciting and interesting about education”. According to him the most important questions to be asked are ones of appropriateness – i.e. whether we should be collecting and using data for a particular aim and the reason we are doing it. Inge Molenaar highlighted that to some extent data-driven services are already taking over tasks from the teacher and with doing that they possibly impact the autonomy of teachers in many different ways. Moreover, she added that issues around privacy and security should also be considered when talking about this subject. Tapani Saarinen, mentioned data bias and privacy concern as well as limiting teachers’ autonomy as the main challenges to be tackled. He also shared that a concern for him as a teacher is that using data would become a “one way street” where students and teachers provide their information without being able to fully understand how this is generated or used.

“A risk that Neil Selwyn sees in data-driven practices is the idea of surveillance pedagogy, where technology is used for tracking, monitoring, and tracing students. In his view, this is especially important risk because schools often do not give options to individuals and students do not have agency to decide what is being tracked and protect themselves and their data. Inge Molenaar highlighted that in her research, students mentioned that they do not necessarily mind the tracking and tracing as long as they keep the autonomy or the possibility to decide who is seeing what.”

What does it mean when the algorithms take decisions that usually the teacher take?

““There’s a fine line between being directed by technology and by being dictated by technology.” Neil Selwyn”

Although teachers, similarly to algorithms, follow a set of rules usually prescribed by the curriculum, Tapani Saarinen thinks that what differentiates the two is the ability of the former to interpret the rules and make exceptions or modifications when needed. Adding to that Inge Molenaar introduced the important distinction between levels of automatization and the relationship between the algorithm and the human teacher. It is important to understand the interactions between the two based on the level of automatization before deciding and discussing the tasks that algorithms are taking over. Neil Selwyn emphasised that when it comes to automated decision making, the tasks that algorithms can execute are very narrow and should be aimed at providing guidance to the teacher who instead should be making decisions. However, it is often hard for teachers to question very expensive and well branded software.

An important element added to the discussion is the question of accountability. Neil Selwyn stressed that teachers often remain accountable for decisions in their classroom, even if those are taken by an algorithm. This accountability is even used to assess teachers, making them vulnerable to situations they do not necessarily fully control. Instead, what Neil Selwyn suggests is that school leadership, policymakers, and/or technology providers should be those accountable.
and responsible for the technology used in schools and the decisions algorithms take.

For this reason, he pinpoints procurement processes as the key point where stakeholders should have important conversations and include teachers but also parents and students to share their thoughts and views.

Marijana Kelentic, Senior Advisor at the Norwegian Directorate for Education and Training shared that in Norway school owners have authority over decisions around digitalisation, but there is a significant gap between municipalities in rural and urban areas when it comes to the awareness and knowledge about privacy and data security. To tackle this, Norway has introduced voluntary online studies for school leaders on digitalisation in an effort to shift the focus from the individual accountability of teachers to the broader responsibility for developing digital maturity in school community. When it comes to the Netherlands, Inge Molenaar shared that schools are fairly independent in making decisions and teachers are involved to a large extent. However, often teachers lack proper information about the function of algorithmic systems because of the absence of transparency on how technology works and how it impacts their decisions. To that end, Inge Molenaar stresses that we need to empower teachers to make informed decisions rather than attributing them with faulty decisions of those systems.

Finland and Sweden follow similar policies when it comes to procurement where the organiser of education, often municipalities or independent schools is responsible for making sure systems are trustworthy and function well. This creates the problem, as Tapani Saarinen highlighted, that organisers often audit the same resources which creates a lot of extra work and costs which could be avoided. Peter Karlberg, Director of Education at the Swedish National Agency for Education raised that this approach, having school organisers as responsible for approving technologies, could also limit teachers’ autonomy to take part in discussions and determine themselves what tools to use.

### What are the necessary steps to ensure teachers possess the right skills and competences to make best use of this data?

Responding to this question Neil Selwyn shared that we are only starting to talk about ‘algorithmic literacy’ and especially what might be termed ‘critical AI literacy’. He outlined some elements teachers need to have in mind when we talk about digital and AI. First step is to be able to recognise when a data-driven automated system is being used. Second step is understanding of how the data-driven automated system works, what Taina Bucher (2017) refers to as algorithmic imaginary. Third and last step is knowing how to work with algorithmic systems. This includes how to make best use of a system but also how to work around a system, knowing when to override an algorithmic decision, or perhaps disrupt or turn the system altogether off. Inge Molenaar concurred with Neil’s analysis and added that in the Netherlands, the National AI and Education Centre is working to bring teachers, business developers, and scientists together for meaningful discussions and to develop a common language and understanding among them. The 10-year programme, funded by the government aims to support this debate in the public sphere as these groups do not naturally exchange.

Although there are several suggestions and offerings for teacher professional development in Europe, Tapani Saarinen emphasised the different levels of expertise and competence among teachers from different areas, schools, or even within same schools. He stressed that there should be a common working culture and a motivating drive instilled in teachers to see the value of acquiring digital skills and prioritising this in their agenda. Moreover, he believes that schools, similarly to other organisations, require strong leadership to support digitalisation and data-driven practices.

“This element of motivation is crucial and that is why teachers who see themselves as not interested in technology should be fully included in these conversations, according to Neil Selwyn. If not, the danger is that in some years from now using technology will be fully embedded into schools and classrooms, and then it will be too late to have critical exchanges with them.

When it comes to some practical ideas, Tapani Saarinen suggested the possibility to pilot different approaches in selected schools before scaling them, while Neil Selwyn suggested having conversations about technology not only in STEM classes, but also in social studies, history, and civics so that teachers and students discuss not only about how technology works but can also explore questions about the politics of data, the role of the private sector and possible alternatives that can exist.
How can we agree on a way forward that harnesses the benefits of AI without compromising students’ privacy and teachers’ agency?

The point, as Neil Selwyn underlined, is not about rejecting AI but rather challenging AI systems to extend what teachers can do in genuinely new and surprising ways we have not yet imagined – not just simply replicating teachers’ work. On the other hand, Inge Molenaar reiterated that the only way forward should be having conversations together and focusing on how AI can augment teachers’ agency instead of taking over their roles; a hybrid intelligence model as she calls it.

“I would challenge anyone involved in AI to try and do something completely extraordinary.” Neil Selwyn

Further reading suggested by the speakers:


Resources suggested during the webinar for future reflection:

Data in Schools - What teachers think (A3 poster) (monash.edu)

Schools, researchers and companies join forces in the National Education Lab AI to work on intelligent technology in education | Radboud University (ru.nl)

About Us — Civics of Technology
https://chat.openai.com/

Homepage | EdSurge Product Index

Projects — European Edtech Alliance (edtecheurope.org)
How can teachers best use and make sense of data in education?

How do teachers know how and what their students learn? Can data support this process?

Teachers know how students learn, according to Olga Viberg, because of the continuous interactions with them in the classroom, the observation of their activities, and the multiple types of measurements they use during their teaching. Guillermo Medrano agreed with this analysis, adding that teachers develop an idea of what and how students learn based on their experience following the students’ academic progress but also getting to know their social situation and environment they find themselves into. This situation relates to Lyndsay Grant’s mention that “teachers are in their classrooms with their students and form opinions and judgments from getting to know those students”.

Olga Viberg argues that data can help this existing process and “raise teachers’ awareness of how students, maybe not learn directly, but how they are engaged in the learning process”. Based on her thinking, the teacher can have an overview of how students or groups of students perform and get some insights related to their learning and the support they might need. This way of using data can, according to Guillermo Medrano, support the teacher but should not be used as the only source of information when assessing a student. For Lyndsay Grant we should be aware of two tracks that can inform teachers of what their students are learning. The first relates to formalised and, often standardised, assessment which creates a lot of data that can track students’ performance. The second one is more holistic and related to the implicit knowledge of teachers knowing their students, as mentioned above. This track does not necessarily create a lot of data that can be tracked or codified in the same way with the first one and can create tension between which tracks educator place more weight on in their in decision making. An important element to keep in mind when talking about use of data in education is, as Olga Viberg put it, for education actors in general to first define the problems they wish such information to help them address. Clearly defining the reason why data is collected, can help schools and teachers understand what type of support they need to visualise and interpret the data in order to take informed decisions. Collaboration is of essence among school actors, but also among them, researchers, and policymakers, so that data is used the best way possible. However, Olga Viberg, also stressed that currently schools use so many different ways and platforms to collect data that analysis and sensemaking is very difficult. When the issue of data ownership comes into play, the complexity only increases because student data ownership involves a long list of actors such as the students themselves, but also their parents, their schools as well as local, regional, or national educational authorities that have access and rights to this data.
Lyndsay Grant shared that, because of the complexity, lack of transparency and limited options in how platforms work with student data, some schools also use more basic software such as spreadsheets to collect and analyse data, where they can have more direct control of the process. In this way they are the ones choosing the data to use and the type of analysis done and are in control of the process.

“Schools and teachers often appreciate the direct control. That’s something that working with spreadsheets gives them. It means that they are not just feeding data into a black box that then comes back to them, but that they can actually choose what functions and analysis to apply to that data.” - Lyndsay Grant

Guillermo Medrano agreed that spreadsheets are used, but in Spain nowadays schools use Learning Management Systems (LMSs) because spreadsheets are not enough for the different types of information they gather. However, from a technical point of view, he argues that he is not aware of a tool that is easy to use and can support the analysis of complex type of information. For Olga Viberg there is relatively little use of learning analytics in compulsory education and data is mainly used to track achievement and performance. In her view “schools are data rich [], but [] information poor so far”.

“Maybe ten years ago there were [only] some teachers that were using learning management systems. But right now, everyone uses one from the three or four [options] that our regional government [in Spain] offers. And every [teacher], has at least a digital repository of tools they use.” - Guillermo Medrano

How can teachers best make sense of the educational data available?

Lyndsay Grant believes that teachers sensemaking of data is closely related to the level of accountability of the education system. In her view, the higher the stakes of accountability targets, the less able the teachers are to use data in an “open-ended way” since the focus is on “producing data that looks good”. The only way, therefore, to allow for schools to engage with data more meaningfully is lowering the stakes of accountability targets in the system itself.

“So when you’re in a very high stakes data-based accountability system, producing the right data becomes an end in and of itself. And that actually stops people saying: What does our data show? Why does our data look like this? What other issues might this be indicating with our curriculum or our school demographics or our style of teaching or our pedagogical approaches?” - Lyndsay Grant

For Olga Viberg it is important to make a distinction between learning analytics and sensemaking. In her view, education actors are still in the process of measuring and deciding what to do with the information that is being collected. Actually making sense of the data is based, according to her, on how people perceive and interpret the information from the context or the environment. This process takes longer time and requires an understanding of a situation in order for teachers to “make an action in the classroom by for example, providing feedback in time or changing maybe pedagogy or, you know, something that would improve students’ conditions for learning”. But in Olga Viberg’s opinion, we now have tools that can help this sense making by providing information that is useful for the teacher. An example of such tools are some Artificial Intelligence (AI) tools that can “screen reading disability and they can tell you how good, how fast the student really reads, how much they comprehend, and with this information a teacher can have a very nice standpoint to start acting”.

“Making sense out of data, I think, requires a kind of journey for the teachers.” - Olga Viberg

In the view of Guillermo Medrano, teachers are aware that there is a lot of data available and being collected during the school time. However, they are also aware that this data is, as he puts it, “a chaos” and in itself have little to say about the learning process or the students learning. What most teachers, therefore, want is to “comb the hair of that data” which can help teachers make sense of the specific situation and act. Having a multitude of different platforms where data is collected does not help this process and can even complicate the ability of the teacher to understand the data they collect. That does not, nevertheless, mean that teachers do not know how to use such LMSs. What is missing, according to Guillermo Medrano is right analysis by the systems of the data to support teachers on how to improve their practice or assist individual students. In Olga Viberg’s view, schools have a lot of data, but often not the right data. That in many cases is a result of the design of LMSs which focus on the sellability of the product rather
than helping connect the different information and thus supporting the improvement of teaching and learning. On the other hand, **Olga Viberg** believes that improving data literacy of teachers is important if we want to increase the understanding of the tools and the trust teachers put in them.

Contrary to these views, **Lyndsay Grant** suggested that “data might not always be the answer to these questions” and that data might not always give the answer to deciding how to improve or adapt to different situations. In her view, teaching is “not just a scientifically and evidence and data informed practice, but actually something that teachers have spent a long-time learning, practicing, informing themselves” where their “judgments are really valid in those moments of how to improve students learning”. Lyndsay believes that “data can help you identify problems but won’t necessarily tell you the solutions”, but that many systems are designed primarily to produce student data in a form that can be easily or simplistically applied:

> “Many of these systems, because they want to be able to tell teachers how to improve, they start to design activities in order to generate data that can then be used to sign-post students to the next particular kind of activity. So, it becomes a way in which the desire to find data, to predict and recommend next steps actually becomes another end in itself, rather than something that is actually being used in a truly analytical sense to understand and to suggest ways forwards.” - Lyndsay Grant

### Educational data and accountability

Responding to a question related to the issue of accountability when using data in education, **Olga Viberg** argued that this is mostly a question for the school leadership and less for teachers who are not involved so much in performance analytics related to the overall school. She also argued that these performance data have undeniably an impact on teachers’ practice and that is the problem. In her words, “learning and teaching should be left to the teachers”. A way to address the issue, according to **Olga Viberg** is to separate the type of data schools collect and the reasons for doing that. She argued that currently there is little empirical research on learning analytics in K12 that is specifically targeting teaching and learning. By separating accountability data from learning data, we can allow teachers “to decide for themselves”. Such an approach would need a “systemic” change of how data are used in schools.

In **Lyndsay Grant**’s view, data in schools and accountability “go hand-in-hand”. Referring to England’s particularly strong reliance on data in a high stakes accountability regime, she shared her concern that many anglophone countries are following suit and increasing data-driven accountability in their school systems. In her view, a way to truly benefit from the use of data is to reduce the negative and “punitive effects on schools for missing targets and actually reframing that in terms of what kinds of support schools might need”. Another approach to system-level accountability, **Lyndsay Grant** argues would be to take a “more statistically literate approach to evaluating education systems”. This means that instead of collecting and publicly reporting school performance data based on entire student population, performance could be analysed through a “sample- and cohort- based assessment so you don’t need to measure every child in every school in every year in order to understand how well your education system is working as a whole”.

### Where should the limits be set when it comes to what type of data is collected or how it is used?

In **Guillermo Medrano**’s view, there is no problem with gathering different data in schools. For him the main issue lies with the security of this collected data. In the past, schools were collecting data in written physical form which had to be properly handled and stored in a secure environment within the school. Today, he argues, schools use so many different third-party platforms to collect and store data and at the end of the day, little is known of how this data is stored or used by the commercial actors that provide these services.

Feeding in the discussion, **Olga Viberg**, mentioned that with the use of commercially available platforms, schools are actually not able to control what type of data is collected. Most often, these platforms track and collect students’ data with limited or in some cases not properly provided informed consent by the students and their parents. **Olga Viberg** also emphasised that students are, in effect, “powerless” since the use of LMSs in schools are many times mandatory and students have no opt-out choice.
She added that although in certain societies, like in Scandinavia, there is a high level of trust and openness to such processes, we should be careful that privacy and security of students are not compromised.

“[We] d[on’t] really do not have much control now unfortunately, but we can maybe put some thinking or practical regulations around how to protect students’ data.” - Olga Viberg

Lyndsay Grant was also of the view that “schools are currently completely not equipped with the technical or legal expertise to fulfil their role as data protectors”. She argues that schools are “data leaky” and it is often hard for schools and teachers to know exactly where their students’ data have gone. On top of that, according to Lyndsay Grant, many school data protection officers have not been trained to fully understand the technical aspects or legislation around third-party data collection and processing and do not always fully understand “how data is being used and shared across third party platforms”, thereby undermining the premise of informed consent by teachers’ and students’. That is especially the case when sensitive information of students such as biometric data, for example facial recognition and fingerprints, are used. These types of data relate more to security and discipline rather than assessment and learning, with a recent controversy in England relating to nationality data collected in schools was shared with government agencies for immigration and citizenship purposes, unrelated to learning and teaching. This is what both Olga Viberg and Lyndsay Grant referred to as “student profiling” and refers to the combination of different types of collected data to create profiles for the students. In that way, even if student data is anonymised, the volume of information is such that it can lead to very detailed profiles that essentially bridge anonymisation. To counter that, Lyndsay Grant believes that legal structures should be put in place where any data collected and used in schools must be justified in the best interest of the child.

“[D]o we think teaching is simply the application of information derived from assessment and other analytics or do we think teaching is something that is situated in a particular context, embodied in particular professionals who have learned and bring their whole wealth of experience to that environment?” - Lyndsay Grant

When it comes to the effect that data use in education has on teacher professionalism, Lyndsay Grant argued that there is a “serious tension on what counts as legitimate educational knowledge”. With that she is referring to the extent to which we trust teachers and their professionalism as education experts. In her research, Lyndsay Grant found that more experienced teachers tend “to be more confident in their own judgments, they have the weight and wealth of experience of being in a classroom [for longer]”, while less experienced teachers might find themselves “relying more on data to justify their knowledge because they are not as confident yet in their own kind of professional experience”. This tension poses the actual question that is what we consider as “teaching”. In her view, data should be seen as a tool in the hands of teachers rather than a tool to control teachers.

“[D]oing teaching as a constant adaptation to the subject that you have this year, [or] to something that happened in a school that affects the whole school or some students” and a teacher has to try do their best given the resources and conditions they have at hand.
From her side, Olga Viberg mentioned that it is important to use data to support teachers be better at their work. In that sense, data can help teachers develop professionally by understanding better their needs. Teachers should be able to use data, and thus be data-literate, in order to support their practice. An issue that Olga Viberg also identified, when replying to a question about the privatisation of teaching as an increasing behaviour, is that there is not a lot of exchange between teachers on the use of data in their practice. There are plenty of communities of exchange for the use of technology in the classroom but only a few projects related to dialogue regarding data-informed teaching practices.

Further reading suggested by the speakers:


Resources suggested during the webinar for future reflection:

Doing Data Differently (shu.ac.uk)
‘Doing data differently’- now open for visitors! | Sheffield Institute of Education (shu.ac.uk)
Home | Lexplore | Systematic reading development
ChatGPT and the role of AI in assessment

Webinar speakers March 2023

Rachel Arthur
Head of Computing at Teach First.

Marco Neves
Expert in Artificial Intelligence in Education, consultant on digital education, and Computer Science Teacher

Mike Sharples
Emeritus Professor of Educational Technology at The Open University, UK.

How does ChatGPT work and what are its key capabilities?

Mike Sharples started the conversation by firstly introducing what is ChatGPT, a term that is buzzing currently on the news, social media, and discussions. In short, ChatGPT is a chatbot or conversation system that allows for users to converse with it in everyday language using text. It has been developed by OpenAI, a company formed in 2015, and was initially released in November 2022. GPT stands for “Generative Pre-trained Transformer” and is based on a large language model. Currently ChatGPT is based on GPT-3 and GPT-4 editions of the program. Mike Sharples mentioned that this language model “is a computer based neural network that’s being trained on billions of web pages, blogs, books and, in essence, on most of the openly available textual content of the web”. That makes it work as “a very highly trained style and text completer”. The difference from the text completers on your phone is that instead of “suggesting just the next few words, it can look back over the previous thousands of words, with the GPT4 version able to generate up to 25,000 words, so it can generate up to 50 pages of text”. The chatbot can “write in any style, in multiple languages, now including minority languages such as Welsh and Catalan”. Mike Sharples also mentioned that the latest edition of the model can “interpret text and images, for example it can answer complex questions that include diagrams”. This can have significant impact on the way students are assessed at school. However, Mike Sharples stressed that ChatGPT and similar technologies also have limitations. One of these issues to be considered is that the data on which the chatbot was trained are not up to date, with most data referring only until September 2021. Another issue is that the technology sometimes “hallucinates”, meaning that it may make up facts.

Mike Sharples explains that this is the case because it is “not a database”. One should, therefore, always check and critically assess its outputs.

“[ChatGPT] is not looking up facts on the Internet [but it is] a creative language tool, a highly trained creative language tool, but it’s not referencing doing a web search. [instead] it is creating plausible language and sometimes the plausible language it creates is incorrect.” - Mike Sharples

These limitations, Mike Sharples highlights, should not deter one to use the tool as it can be “hugely empowering”. This empowerment will only augment as companies such as Microsoft, which has invested substantially in OpenAI and ChatGPT and is embedding this technology in its Office Suite, are creating new opportunities and possibilities for users.

“We can look at ChatGPT as a student in the classroom that always answer. You ask him a question and he always give you an answer. The problem is [that] some of the times [this] is not the most accurate answer.” - Marco Neves

Marco Neves in his introduction mentioned that Artificial Intelligence (AI) is “not something new”, but technology such as ChatGPT make it even more obvious and tangible for most of us. However, AI and what it brings is, according to him, a “huge challenge” for teachers especially and
we should help them understand the technology and the impact it can have in their work. In Marco Neves view, we cannot address AI as “just one more tool” and that is why the focus on AI Literacy is very important.

“Artificial Intelligence is not something new, we should not forget that. [But] ChatGPT was like the elephant in the room. The elephant was entering [the room] a long time ago, but we didn’t look at it and now we are completely surprised because we somehow didn’t do the homework.” - Marco Neves

One of the main differences of ChatGPT to previous technology tools, according to Marco Neves, is that Generative AI systems like ChatGPT “use text, images, videos that already exist to create something new”. The output is thus completely new. This begs the questions of whether teachers are able to identify whether a student’s work is produced by themselves or by a chatbot. For Marco Neves if a question given to students can be answered correctly by a chatbot, then “the problem is not the student or ChatGPT”. Instead, the real question one should ask relates to what type of questions students are tasked to answer what the purpose of it is.

Marco Neves believes that education systems are facing a huge challenge when it comes to the educational models and strategies used today. For him, “it is critical that teachers start to understand what is happening, they learn about these technologies, and reflect on the ways they are preparing their students”. Companies already start integrating the technology of ChatGPT in their systems, often in education contexts, allowing students to interact with chat-assistants that “can interact” and “have a conversation” with the student. Ultimately, for Marco Neves, it is important to consider what the role of the teachers is and what profiles of teachers are needed today.

From her side, Rachel Arthur stressed that as educators “we’ve got to prepare our peoples for the reality and the future reality that they will be facing in the world of work”.

Rachel Arthur believes that since teachers are meant to prepare students for “this brave new world”, “simply banning it [ChatGPT] and running away from it, isn’t the answer”. However, such tools raise concerns about biases and even racist outputs, as well as questions about equal access to quality education.

“I have a big fear around the digital divide. We already have pupils that don’t have access to computers at home, and therefore they won’t have access to the same AI tools that other people from more privileged backgrounds would have. So, is this tool [ChatGPT] going to make that divide worse or does it have the power to make it better?” - Rachel Arthur

How does technology like ChatGPT impact educational assessment and teachers’ work?

When it comes to the impact of AI on teachers’ work and the necessary preparation and upskilling of educator’s competences, Rachel Arthur shared that in her work the main focus is not on the tool but on addressing the main challenges that come with the specific technology. In her view, in an era where most teachers have had by now -at least- some exposure to technology and digital tools, and where available tools change and evolve rapidly, it is better to focus on the “bigger picture” and how best to prepare individuals to best address the needs of the future. This bigger picture is also crucial to consider when trying to answer the question of what the purpose of education in the 21st century really is.

“[This] brings up a bigger question of what the purpose of education is. So, are we preparing people to sit exams? Are we preparing people to remember things, to regurgitate things, to follow mark schemes, or are we preparing people to be the best that they can be, the best digital citizens of the world of tomorrow?” - Rachel Arthur

Rachel Arthur believes that the most important objective related to the future of education is to train teacher to understand the challenges that AI in education bring and to prepare them to address them with confidence. That is connected to the overall idea of teaching students how to think critically and equipping teachers with the tools to support them understand and identify the risks that AI can bring such as extensive biases and incorrect or made-up information (connected to the hallucination phenomenon).

In Rachel Arthur view, to address the impact of AI on educators’ work, there is need to make sure they are well-prepared, trained, and literate on the topic of AI so that they can best harness the benefits this technology offers.
Focusing on the impact that Generative AI tools like ChatGPT can have on students’ assessment, Mike Sharples underlined that he does not believe that banning such technology is the solution. The technology is improving at increasing speed, and it will become harder and harder to identify whether a piece of work is generated by a human or by a machine. Moreover, Mike Sharples believes that banning the technology would only exacerbate the existing inequalities among students.

In his view, confident students would continue to use such tools, perhaps as a means to formulate a first draft that they can revise and adjust according to their style and ideas. On the other hand, less confident students would avoid using the technology altogether as they would be worried that they would be caught. Instead, Mike Sharples proposed a plan that looks to address the issue of assessment in a short, medium, and long term. In the short term, what he proposes is the introduction of more “oral or invigilated examinations”, especially when it comes to final assessments. This should be coupled with what Rachel mentioned about professional development of teachers with an emphasis on AI literacy which is different from digital or computer literacy. AI literacy, Mike Sharples mentioned, focused on the way AI systems work, their benefits, opportunities, but also limitations. When it comes to the medium term, Mike Sharples suggests that adaptation of assessment approaches is needed. That could mean more project work for students which limit the value of using AI since the chatbot is unaware of the specifics of the project work. Finally, in the longer term, Mike Sharples proposes the introduction of new, more dynamic forms of assessment (formative assessment) to empower students support their learning.

Echoing the other two speakers, Marco Neves reiterated the need to support teachers in this very crucial moment in human history. Teachers, he argues, are already struggling with many different things in their everyday work. It is therefore paramount that they are empowered with proper professional development opportunities and training to address the new challenges that arise. Marco Neves also mentioned that when it comes to the challenges AI brings, the use of the new tools is very low in terms of importance. What is the greatest challenge that AI technology has created is the antagonism in what means to be human.

“What never faced something like this in terms of our history as humans. For the first time we are being challenged in our last exclusivities in terms of [what makes us] humans.” - Marco Neves

This new challenge and the speed that new AI tools enter the public sphere and become broadly available create the urgent need to focus on educating individuals, and especially teachers about AI literacy in order to avoid further increasing the existing digital divide. Marco Neves believes that empowering teachers with AI related skills is fundamental to prepare students for a world where interacting with such technology will be critical. For this reason, Marco Neves makes a plea for European countries to work together and develop a Framework on AI Literacy which can act as a guide for professional development programmes of teachers.

What should the focus be when covering such technology during teacher training?

Asked about their thoughts on the focus that teacher trainings should have when covering generative AI and AI in general, Rachel Arthur shared some of the key areas her programme at TeachFirst is covering. The five areas covered relate firstly to what AI and AI in education is and secondly to how teachers can use AI in lessons, exploring different tools that can be used, focusing on adaptive technology and how it can support learners with special educational needs. The second area covers user agreements, rules and regulations and parental permissions because, as Rachel Arthur mentioned, there is a lot of confusion about whether such tools can even be used in the classroom with underaged students. The third area focuses on looking at how students can use AI to support them, to continue their learning at home and on some of the issues related to the technology’s limitations such as the hallucination of fact and fiction and the biases that exist. The fourth area looks at how teachers can use AI to manage workload without getting themselves in any trouble, the ethics of using AI to generate lesson contents, using AI for assessment purposes. The fifth area relates to incorporating AI in the curriculum as a subject topic.

In his intervention, Mike Sharples shared an open repository of resources curated by Anna Mills, a pioneer in the use of ChatGPT and AI language models in education. Mike suggested that the repository can act as a guide to anyone developing professional development guidelines or courses for teachers around this topic.

From his side, Marco Neves shared the main elements the programme he has developed for teacher professional development in Portugal entails.

---

The 15 hours long training starts with an exchange among the participants and their role as teachers. The programme continues with reflection on their practices and assessing their beliefs on the role of education today. The programme, Marco Neves argues, aims to explore different tools and strategies that teachers can employ to experiment with and enhance their practice. It also aims to help teachers rethink of the way they see their role and reiterate their position as central to the schooling process.

### Can ChatGPT support student agency, creativity, and personalised learning?

AI and tools that use generative language models like ChatGPT have the possibility, Mike Sharples argues, to support innovative and new ways of teaching and learning. Students can be empowered to use such tools as a support to complete and enhance their work. Mike Sharples believes that it is a matter of properly preparing educators and students to best use these tools for their own benefit. AI development will create more tools and options for students to use. Teachers should therefore, with the support of the school and educational authorities, develop proper strategies related to dynamic assessment in order to best harness the possibilities that AI offers. In his view, Mike Sharples, thinks that if students are willing to share aspects of their interaction with AI tools, using these data can prove very beneficial for teachers to better support them and provide the necessary targeted guidance and resources. He believes that instead of seeing schooling in opposition to the AI tools that students use to empower their learning, it is important to look for congruence and ways in which they symbiotically coexist and support each other.

“"I see a synergy, a congruence between students who are supported by these [AI] guides, mentors, and support tools and teachers who have the ability to see where students are in terms of their learning and their knowledge and what sorts of support they might need to progress further."” - Mike Sharples

Adding to the conversation, Rachel Arthur totally agreed with what Mike mentioned and added that AI tools provide the possibility of freeing up teachers’ time to be able to personalise students’ learning without having to deliver the same content to all. In that way students can decide for example the article they would like to read or progress at their own pace when learning to code with an AI assistant providing instant feedback. When it comes to home learning, Rachel Arthur argued that AI chatbots can help students have equal access to quality support, provided that the tools are able to provide reliable and valid information. However, the teacher should always, Rachel Arthur states, remain the expert in the room with AI tools acting only as support and not as replacement. Teachers should not be scared about losing their jobs because AI “cannot beat that human interaction […] no matter how sophisticated these systems become”.

“We’ve got to remember that the teacher is the expert in the room and […] that pedagogy is so important because these bots are algorithms. They’re not humans. They don’t know your pupils like you do. They don’t know the stories of every single child.” - Rachel Arthur

Marco Neves also reiterated that we should not forget that “learning is mainly a social interaction, and these systems are not even close to having any kind of [such] capabilities from an emotional point of view”. AI tools use language but they do not actually understand it or make sense of it. However, Marco Neves argued that such tools can support producing “synthetic knowledge” which can help enhance students’ creativity and teachers’ work.

### What are the ethical considerations when using ChatGPT in education and how can they be addressed?

Addressing some of the ethical considerations related to generative AI models Mike Sharples highlighted that it is often argued that these tools do not have human experience, but they might develop, over time, experience of their own world – the internet. The reason why companies such as OpenAI have not allowed this to happen is because they are worried of the repercussions this might have.

These repercussions related to the ways AI systems could interact with the live data of the internet, interpret specific situations, and create outputs that might cause harm. However, Mike Sharples, thinks this will eventually happen and teachers should be ready to face such reality.
On the other hand, Mike Sharples also mentioned that educators and education systems should be well aware that AI systems have no empathy nor care about anything or anyone. They might show such signs but that does not mean that the system has any emotional feeling towards the user of the system of humanity overall.

What is worrying, therefore, Mike Sharples explained is that as these systems become more and more widespread, teachers would have the task to distinguish human from machine interaction and prepare students to critically assess this interaction.

“[As AI systems become] more and more part of our everyday lives, things that we converse with routinely, they will give the pretence, the simulation of caring. They will seem to be able to answer your problems, to be able to sympathise with you, to support you, but they will do it, if you like, in an algorithmic way. And I think as teachers we need to show that there is something different about humans, that we care because we are humans and because we have a shared understanding, a shared empathy. That is likely to be a new role of teachers: to show that we as teachers care for our students, we don’t just interact with them.” - Mike Sharples

Marco Neves stressed that we should also look at these AI systems from a cultural perspective. For example, he mentioned that if you go to “countries like Japan, South Korea and so on, the way they interact with these systems and the way that they easily integrate them not only in education but even in other fields is completely different from us in terms of Europe”. Marco Neves also mentioned that the way we train teachers should be different than in the past, because AI is a completely different technology.

“We cannot train our teachers like [they were] trained for [using] the Interactive Whiteboard or something like this. This [technology] is more and more complex, so if we keep doing the same thing as we’ve been doing so far, we’re going to have a lot of very complex problems in the future.” - Marco Neves

Another important point according to Marco Neves is to recognise that the data based on which AI technologies are trained have been created by humans. This means that the data reflect all the biases and stereotypes that exist in our societies which feed into algorithms that train AI systems. However, it is more important to understand that although the systems are relying on human data for now, in the “near future they will be trained in the data that they create by themselves, creating […] a snake [that] is biting its own tail”. Marco Neves underlines the need for specific action to be taken urgently by governments and education institutions in order to address AI divide in its birth and not allow another huge gap exacerbate inequalities among individuals. In his view, “companies are focused on profit” and we already see many of them firing their AI Ethics personnel or dissolve whole teams dealing with ethical AI development. For this reason, Marco Neves argues that we need to engage both teachers and students to prepare for the future that is already here.

Continuing his intervention Mike Sharples distinguished two big ethical issues: inherent bias in the data and ownership of the systems. On the one hand, Mike Sharples argues, in connection to what Marco mentioned as well, that AI language tools “have been trained on billions of words, and ironically, the more they are trained, the more they reflect the bias of the data on which they’re trained”. This way, these systems reflect all the cultural, racist, misogynistic biases that are inherent in the data that they were trained on. Of course, Mike Sharples recognises that some companies go a long way to remove such biases using humans to check and remove them from the models. Some companies, however, might not consider such a process important or might lack resources to evaluate large amounts of data and outputs and label them as biased or not. On the other hand, Mike Sharples reasons that today most of these AI systems are owned by big technology players who can afford huge investments. Thus, it is these companies that have the power and influence to control how the algorithms look, what data is used to train them, and what is considered as bias in the data. Moreover, big technology players are mostly coming from the US and China. Mike Sharples raised then the question of what the implications could look like for individuals from outside these countries when it comes to languages available, cultural understanding, norms, and conditions. This question is especially important when considered under the lens of using such tools in education and for learning.
From Rachel Arthur’s perspective, perpetuating biases is an important consideration we should keep in mind. Gender, racial, and cultural characteristics and stereotypes feed algorithms that then are used in products that mirror the exact biases society is trying to change.

“The fact that if you ask Alexa or Google Home to respond to you, it responds in a woman’s voice is that increasing the subordination of females by it being a female voice automatically? Why aren’t there male voices? [Or] the fact that facial recognition software doesn’t recognise Michelle Obama or Serena Williams […] are problems that are baked into the system in the way that we design it.” - Rachel Arthur

Rachel Arthur believes that unless more women, people from different cultural, linguistic, and ethnic backgrounds join the tech world, writing code, and making decisions about AI systems, these biases will continue to exist and will only increase. For Rachel Arthur, it is “the people that are designing this software [that] have the power to design the guardrails” and unfortunately this power currently sits with a few big corporations. In her view, all people should have access to technology in equal terms, otherwise the digital and AI divides will only widen. Rachel Arthur thinks that “if people don’t have access to this technology, then they can’t engage in the same way that others can”.

Further reading suggested by the speakers:

Arthur, R. (2023) “From Chalkboards to Chatbots,” rachelarthurwrites, 2 February. Available at: https://rachelarthurwrites.wordpress.com/2023/02/02/from-chalkboards-to-chatbots/


Resources suggested by participants during the webinar for future reflection:

How Will Schools Respond to the A.I. Revolution? #chatgpt #artificialintelligence - YouTube

Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators | European Education Area (europa.eu)

AI Text Generators: Sources to Stimulate Discussion Among Teachers - Google Docs

InVID Verification Plugin - InVID project (invid-project.eu)

Artificial intelligence in IB assessment and education: a crisis or an opportunity? | IB Community Blog (ibo.org)

[2303.10130] GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models (arxiv.org)

Sci-Fi Short Film “Hard Reset” | DUST - YouTube

Marco Neves on Twitter: “The #ArtificialIntelligence (AI) Divide will produce much more profound negative impacts and inequities than the previous Digital Divide. We must guarantee that it will not happen. Only through #education we will achieve it. https://t.co/ExterYQRAu” / Twitter

The Nadella and Pichai Face Off in AI (analyticsindiamag.com)

Ethics of AI in Education: Towards a Community-Wide Framework | SpringerLink
Resources shared by participants in the registration form:

- Artificial Intelligence in education: challenges and opportunities | ESEP (europa.eu)
- The use of AI in learning and teaching: first steps and feedback | ESEP (europa.eu)
- Results of the teachers’ survey on Artificial Intelligence help European Commission to develop ethical guidelines | ESEP (europa.eu)
- The impact of social robots on children’s development: What science says | ESEP (europa.eu)
- Promoting children’s fundamental rights in the era of Artificial Intelligence | ESEP (europa.eu)
- EU policies to support and govern the use of artificial intelligence in education | ESEP (europa.eu)
- Course: Webinar: Using artificial intelligence to support teaching and learning in schools (europa.eu)
- Listen to the new Teachers of Europe podcast – giving a voice to inspiring teachers from across Europe | ESEP (europa.eu)
- Informatics education at school in Europe | Eurydice (europa.eu)
- Digital Education at School in Europe | Eurydice (europa.eu)
- What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education | Smart Learning Environments | Full Text (springeropen.com)
(How) is using data for learning ethical?

Webinar speakers May 2023

Duuk Baten
Responsible AI advisor at SURF, the Dutch National Research and Education Network (NREN)

Wayne Holmes
(PhD, University of Oxford), Associate Professor in the UCL Knowledge Lab at University College London

Sonia Livingstone
FBA, OBE, Professor of Social Psychology, London School of Economics and Political Science

What do we mean by ‘ethical’ data in education?

Addressing the question, an interesting linguistic observation was made by Duuk Baten who challenged the idea that there is such thing as “ethical” data or “ethical” Artificial Intelligence (AI). Instead, Duuk suggested that the better way to phrase it is to talk about ethics of data or data ethics. In this way, we avoid giving agency to data or an AI system. Building on this argument, Wayne Holmes emphasised the importance of understanding that “words have life of their own” when chosen to describe things, but also that “once they’re being used, they can start to limit the way that we’re thinking about the thing in question”. He argued that if researchers 60 years ago had given, for example, the name “advanced cybernetics” instead of AI to the systems they were developing, we might have not been debating whether these systems have sentience or intelligence. Therefore, when talking about ethics, it is important to understand that it is humans who make choices and have agency rather than the machines or the information collected and used.

Going back to providing a definition of ethical data or ethics of data, Wayne Holmes suggested that we refer to “something that by design and in practice protects learners’, teachers’, and other stakeholders’ in education human rights”. That includes, in Wayne’s view, areas around agency, autonomy, safety, dignity, and protection from commercial exploitation. However, Wayne Holmes believes that many issues appear with the data themselves throughout their life cycle. The issues begin with the way data is collected, including questions of who collects it, for which purposes, and the choice people have in consenting for their data being collected.

A second issue where ethics come into play relates to storage and ownership of data. In particular the safety of the storage, the accessibility to data sets, and the anonymity that pertains it. Important also for Wayne Holmes is the question of ownership of the data and the ability for the owners to use the information for future purposes, the usefulness of which cannot be yet determined. Next issue concerns the analysis and interpretation of the data, the impact of the outcomes to the data subjects, and the way the results will be used – e.g., to support learning or commercial objectives. Finally, important issues to consider are the fair and equitable treatment of data throughout their life cycle.

Feeding to the discussion, Sonia Livingstone, connected the point about human rights to more particularly children’s rights and proposed the United Nations Convention on the Rights of the Child as the guiding framework when thinking about crucial issues affecting young individuals. She also added to the questions that need to be considered when we talk about ethics of data, underlying the need to be clear about “to whom should data use be transparent, to whom should terms and conditions be comprehensible, and how does one claim remedy against wrongs?”. These questions are even more challenging when children are the data subjects, because it is very hard to imagine what the long-term implications could be for them in the future. Sonia Livingstone also underlined the duty we have to care and to use data to benefit children’s education.
For this reason, data should be used in a way that respects their privacy and allows their voices to be heard. Moreover, she pointed out that although we live in a world of public-private arrangements, we should be wary of the delicate boundaries between use of data by businesses for mandated purposes and commercial exploitation of students’ data.

“The overarching questions are how do we make a risk-benefit calculation when we use data and [how] do we know enough about the risks of when data use can go wrong and become unethical? How do we weigh that against the potential benefits? And do we have evidence of the benefits of using education data?” – Sonia Livingstone

Duuk Baten brought up the element of “justified choice” when it comes to the use of data that is part of the European Commission Ethical guidelines on the use of artificial intelligence and data in teaching and learning for educators. In his view, it is important to consider the ‘why’ and ‘how’ someone is using educational data and how are the systems used set up. Duuk Baten also brought up the importance of the ecosystem of stakeholders that are involved in making sure of the ethical use of data. This ecosystem consists of the governments that develop the policies and regulations, the designers of the systems, the actors that sell the systems, those that procure them, the project leaders and IT administrators that manage them, and the teachers that use and implement them. This, Duuk Baten, argued is a “super complex and technical system” and every step has part of the responsibility. In such a complex environment, it is natural that there exists a “diversity of thoughts”, but what is important, according to Duuk Baten, is “weighing on values, weighing on the things that we find important to make those right choices”.

Who is responsible that data in learning is used ethically?

In the question of responsibility when it comes to ethical use of data, Duuk Baten, underlined his previous mention to the complex system and the interplay between different actors in the system. In his mind, it is important for public sector education institutions to accept and get more responsibility controlling and owning education data, reversing the current situation where “a lot of responsibility is actually sort of given away or deferred a bit to avoid taking the responsibility by either outsourcing or hiring consultants or taking certain EdTech providers on board”. In this way, Duuk Baten believes education institutions will have more control over their data and will be able to better understand and analyse them for the improvement of education. Of course, he argues that more control means also more risk, but this is the way to “enable us to be in the driving seat [determining] the future of education instead of becoming more and more passive and allowing more and more of these tools to be developed outside of our school walls”. To achieve this, it is important to provide the right support and incentives for schools to be able to take this responsibility.

In this sort of complex social technical system, there is a whole lot of different people in, responsible and involved, and especially in education you have this public-private interaction, there is an element of also taking responsibility and giving responsibility away – Duuk Baten

Sonia Livingstone from her side, emphasised that those that have the responsibility to act are the governments and educational authorities of each system. However, what is often the case is that “they don’t necessarily have the power to act if they don’t own, or indeed don’t understand the data that is being collected and the uses to which it would be put”. On the other hand, Sonia Livingstone understands that “in the world of complex cybernetics, it would be beyond the capacity of the education system perhaps to generate the insights that are potential from the data” and therefore, “the commercial sector, the EdTech sector is needed both for the provision [of technology] within the schools, but also for the behind-the-scenes expert analysis [of data]”. That creates a delicate balance between the public and the private sector which is important to focus on “harnessing the insights […] and bringing those back to the classroom”. What is important for Sonia Livingstone is that we can reflect upon the questions: “Do children understand the data being collected and how it’s used? What is the relationship of public-private partnership in the world of analytics and data analysis? How does it [data use] come back fairly and does it come back sufficiently fairly that the collective good is served?”. For her, there seems to be a delicate balance between the responsibility of governments and education authorities to legislate and regulate, and the responsibility of each school to make decisions about the technology they are using. However, schools seldom have the “enormous specialist expertise” needed which is, according to Sonia Livingstone, “unrealistic” as it would require significant amount of technical, legal, and technological upskilling of schools which are, at the same time, facing teacher shortages and many other pressing issues.
Wayne Holmes brought a slightly different dimension to the discussion, questioning not the actor responsible for ethical use of data, but the data use for learning altogether. He argued that “the so-called data for learning” is primarily being collected by commercial actors via software provided to schools for profit. For this data collection to be considered ethical, Wayne Holmes believes it is fundamental that it supports students’ learning. If not, then he argues that “however else it might be ethical, the collection of that data just cannot be ethical, because at the very least, we’re wasting the time of our young people to have to engage with these tools that actually don’t do what they say”.

“\textit{The key question about the ethics of data collected in learning is whether that data actually does what they claim it does, which is to support the learning of young people in otherwise ethical ways.}” – \textbf{Wayne Holmes}

Wayne Holmes further argued that “there is so little independent evidence at scale that these tools do anything in the classroom of any benefit to anyone”. Although there are some small-scale academic studies, when comparing with studies done in the medical field to test the effectiveness of medication, there is not enough evidence to prove that particular technologies do what they promise to achieve. Furthering his criticism, Wayne Holmes disputed the promise of learning analytics as a way to understand students’ learning. He reasoned that “the problem with the learning analytics research space is [that] so much effort is put into checking that the prediction that they [researchers] are drawing from their analysis of the data that they have collected is an accurate prediction”, making it a “technical discussion”. In that sense, Wayne Holmes continued, “there’s a huge disconnect between the prediction and the actual impact of that prediction on the individual children and their learning trajectory, on their way of existing in the classroom, and the relationship between the student and the teacher and the student and other students in the classroom”. In essence, he also added academics in the equation of responsibility for using data ethically.

\textbf{In times where ‘open data’ for research is advocated across Europe, can education data be ‘open’ for public use?}

Peter Karlberg, Director of Education at the Swedish National Agency for Education posed the question of ‘open data’ in education and the concern he has around educational authorities publishing educational data for public use. Addressing this question and concern, Sonia Livingstone highlighted that the topic is very important as although open data refers often to anonymised data, in education this can be reidentified as high numbers of sensitive data is collected within this context that can be put together and create very detailed profiles. Sonia Livingstone also raised the issue that in education, companies collect so many details that one can end up with micro-data like keypresses and time one takes to complete an activity. In an area where consent is already questionable, she believes it would be unethical to make students’ data public as this would allow private entities to be the ones to benefit most by using it for profit.

Duuk Baten drew a distinction between data collected in public space, which makes sense to be made public and further used for other public purposes, and sensitive learning data that could be used for unknown purposes. What Duuk Baten suggested is the idea of “federated access” where access to data is given to different groups of individuals based on the sensitivity they carry. In that way, researchers can still have some access to a dataset to conduct research, making sure specific checks and balances are in place.

“We realised that openness is not ideal because openness also means openness to everybody and access to the strong players which are usually these commercial players to keep using it in ways that we might not have wanted to.” – \textbf{Duuk Baten}

Wayne Holmes suggested caution to the context certain words are used for the meaning they carry. The word “open”, he argued, is a positive word when it comes to for example open educational resources (OER), but it has a less positive connotation when it comes to open educational data. Wayne Holmes, similarly to the others, argued that he is in favour of open data about infrastructure for example, but sharing individuals’ data -even anonymised- can be “hugely problematic”. This “Silicon Valley push” carries the “notion that actually, if we share data, we make data open for everybody and everything and then the world will be a better place”. Wayne Holmes thinks that “the principle of open is good” but one needs to be aware of “the way that it’s implemented and to whom it is imposed”.

\textbf{Meaningful and ethical use of data in schools}
In his reply to the question, Wayne Holmes argued that when we talk about technology or data in education, we often forget about the ethics of education itself. This is something that has “been questioned and debated over 2000 years and probably before that”. That is why, when we talk about technology in education, it is important to look at the pedagogical soundness of a particular tool, whether it “actually does what it claims to do”, and the evidence that support this. Although Wayne Holmes understands that “one person’s evidence is somebody else’s hearsay”, he argued that the bottom line is to know whether “these technologies enhance teaching, they empower teachers or they disempower teachers, do they give agency to students?”.

Another important consideration relates to the development of technology for education. Wayne Holmes shared that it is known that “most developments of technology for use in education mostly aim at the average student because what they’re [companies] trying to develop is tough”. This way, he argued, EdTech developers “aim at the middle of the bell curve” of the distribution of students, but then transferring the application to those at the edges of the bell curse (the students with bigger learning challenges) is very difficult. What Wayne Holmes suggested is that “we should start by working with those who have the biggest inclusion needs, and we should be making education appropriate for them, not trying to force them to be appropriate for the education that we have”. In this way, he continued, “it becomes much easier to transfer what you’ve learned from those at the edges of the bell curve to those in the middle”, considering “equity and inclusion from the very beginning, not as a bolt-on at the end”. Adding to that, Wayne Holmes also highlighted the importance of safety and wellbeing of students. He argued that it is important to make sure “technologies that are being developed are, before they’re put into use, actually safe by design”. Companies should also “prioritise wellbeing of the students by design and avoid making young people accidentally coming in contact with inappropriate materials”. All these, Wayne Holmes stressed should “be done by design, not as an afterthought”. Finally, he also underlined the importance of developing and enhancing digital citizenship of everyone, including young people to help them “self-actualise, do it safely, maintain their wellbeing, and help them contribute to society”.

Sonia Livingstone argued that, when answering this question, one should consider whether to question the technology itself or the company that develops it. Moreover, one should distinguish between “winner takes all” companies that are giants in the education market and “hoover” students’ data, and smaller players that are more targeted to specific areas of the curriculum or education provision. She also brought up the different kinds of technology use in education. On the one hand, there is the technology use for developing skills, needed facts, some kind of personalisation, rote learning, and assessment. On the other hand, Sonia Livingstone argued, there are affordances that technology can provide but are “enormously underused”. Those relate to opportunities for networking among students, collaborative and creative learning, tinkering and making, and to support “child-led learning”. However, Sonia Livingstone admitted that pedagogy and use of data is a “serious political question”, and different actors have different views on what beneficial use of technology in schools is.
Ownership and development of technology tools and the primary beneficiaries of its application. Schools, Duuk Baten continued, have certain control over what is happening within their walls, but what takes place outside can also influence the way they work. Students use technology not only in school, but much more outside of it, and this can affect the way they see the purpose of their education.

“Schools cannot be the only ones bearing the responsibility for this [technology use] because students are also autonomous beings. They’ll use interesting stuff, shiny new tools, and that influences their perception of the way they see education, the way they learn from education.” – Duuk Baten

In defining what is ethical for technology, Duuk Baten suggested three elements to be taken into account. First element relates to legal and regulatory frameworks that should put safeguards and should be developed with human and children’s rights in mind. Second element to be considered is the values that drive the different approaches to education and act as a “common language”. Those, according to Duuk Baten, can be formed on a national, sectorial, or even school level and should encompass what is considered important. These values would then be part of design and procurement systems, funding mechanisms, and data infrastructure. The final element relates to “ethics or responsibility as a practice” where every actor and professional must act responsibly, “not hype things that we don’t know about enough, try to be honest about what we’re doing, try to enable our own virtuous behaviour as individuals”. Duuk Baten argued that by acknowledging that the “reality is dirty, it’s gritty, and it’s difficult” and that “there are no easy right answers” we can “go for that technology worth wanting”.

Sonia Livingstone echoing the suggestions Duuk made, mentioned the work that the Digital Futures Commission is doing to develop a set of 11 principles for designing digital products for children. She, however, also underlined the need that educators have for more “pragmatic requirements” that can help them when choosing technology to use in their lessons which fulfills some minimum standards. The Digital Futures Commission is therefore also developing a Blueprint for certification, which can allow schools to assess technology based on some minimum criteria.

Setting minimum standards and certification, Sonia Livingstone argued, should be the work of education authorities which would help guide schools and at the same time, provide certainty for the industry on the regulations they need to abide by.

Adding to the conversation, Wayne Holmes also shared the work that the Council of Europe is doing on working towards a Convention that focuses on Teaching with AI and data, looking at “how we can use technologies effectively and innovatively in education settings”. It will act as a Blueprint or guidance for governments to set, as Sonia mentioned, some minimum standards that can be adapted according to particular contexts and circumstances. With this opportunity, Wayne Holmes also raised the often-quoted promise from EdTech providers about “automating and speeding up what they think we do in education”. He claimed that “the difficulty of doing this kind of work is the problem-based versus solution-based approaches to developing education technologies, and almost all education technologies are solution based”.

Wayne Holmes argued that EdTech companies should start asking educators what the problems are they want technology to help them with, rather than the other way around. In his view, “the Edtech companies [tend to] go for the low hanging fruit, what’s the easiest way to do something and actually don’t really explore the possibilities” which can help us address the question of “what do we want from education” and “what is its purpose”. However, as mentioned above, Wayne Holmes concedes that “we don’t have agreement about that” and therefore it is “very difficult to understand what [the role of] technology should be in education and what we perceive as being ethical uses of technology”.

How can schools, teachers, and parents say yes or no to something that they do not really understand?

Addressing this question, Sonia Livingstone suggested that “no one should ever say yes to anything they don’t understand, and that would not be informed consent if they did so - it would be invalid”. Sonia Livingstone underlined that it is paramount that schools, teachers, parents, and students should be able to understand to what they are consenting to. In her view, “if we don’t understand we cannot consent, so it means something is wrong with the architecture of
choice”. In this sense, consenting to something should be a genuine choice. Coming back to minimum standards, Sonia Livingstone offered the idea of a sandwich when it comes to consent. The bottom part refers to certification and minimum standards set by education authorities, while the top part refers to principles, values, and children’s rights. In-between those, Sonia Livingstone suggested should be the ecosystem that “sets out possible visions that advises on procurement that can offer independent evaluations of what’s working”. This part should be “a kind of a moving space because people are experimenting and trying things out”, where also ideas about education are shaped and changed over time with the advance of technical and empirical knowledge.

Wayne Holmes built on Sonia’s arguments and added that although he considers informed consent as essential, the license approach consent that most technology products currently use does not work. In his view, teachers are already overwhelmed with their duties and thus responsibility should be with the “ministries to ensure that teachers have good quality professional development in this domain”. Teachers, Wayne Holmes argued, should be given real space for professional development – this is the only way to consider it an opportunity for them. He also mentioned that providers of technology used in education should be “open and transparent in genuine ways”.

From his side, Duuk Baten argued that “responsibilities are so stretched throughout this complex ecosystem of decisions being made and there is an interesting interaction happening where responsibility needs to come with an ability”. This ability refers to think and understand what one is reading, for example in the Terms and Conditions of a particular product. Finally, Duuk Baten also questioned the ability of students to really consent within the framework of the school, due to its hierarchical relationship between the school and the student.

Further reading suggested by the speakers:

- Other publications the speakers shared: Data4Learning. (How) is using data for learning ethical. Resources
Resources suggested by participants during the webinar for future reflection:

- Weighing Values (kennisnet.nl)
- Microsoft Word - Problems with data governance in UK schools - the cases of Google Classroom and ClassDojo.docx (digitalfuturescommission.org.uk)
- Update afspraken en gesprekken SURF en Google | SURF.nl
- Child Rights by Design - 5Rights Foundation & DFC (digitalfuturescommission.org.uk)
- A Blueprint for Education Data (digitalfuturescommission.org.uk)
- The Ethics of Artificial Intelligence in Education: Practices, Challenges (e-bronnen.be)
Governance processes for a robust educational data ecosystem

Opening the discussion, Melanie Ehren addressed the first question by exploring the meaning of the word ecosystem when discussed in the field of education data. She used a “broader conceptual framework” to break down its structure into a micro, meso, and macro level with each focusing on specific elements related to the ecosystem. On a micro level, Melanie Ehren identified “teachers using data to inform their teaching and instruction”, on a meso level she included “schools who aggregated student achievement data and other types of data to think about ways to develop and improve the school”, and on a macro level she added “data used by inspection agencies, for example, or ministries of education to inform their evidence, informed policy making”. To talk about a robust ecosystem, Melanie Ehren considers important that one looks on all the different levels and check the “accuracy of data, but also the extent to which data represents different types of student groups, including those who are vulnerable or [part of] minority groups within the education system, [as well as] how data is used to benefit student learning”.

Kruakae Pothong, coming from a more child-rights perspective, focused on the need for effective regulation and market incentives to a robust ecosystem that processes data lawfully, respecting the rights of the children, and is of benefit to their learning. She also raised the lack of transparency that often surrounds companies’ privacy terms and conditions and the complexity this adds to the ecosystem. This complexity is further heightened with the particular role that schools play as “intermediaries between EdTech providers and children”. In this situation, Kruakae Pothong identified a “power imbalance between EdTech, who have the technical control over the data being processed and the schools with the responsibility to ensure that the EdTech tools they use only process the data strictly according to [schools’ defined purposes]”.

Barbara Wasson, from her side, identified the ecosystem as a set of different actors and structures that include students, teachers, and parents, technology providers, the current laws and regulations, and institutions like schools and universities. This “complex system” of actors brings about, according to Barbara Wasson, different technical, regulatory, and competence issues that need to be addressed.

When asked about the role of the EdTech providers specifically in the ecosystem, Melanie Ehren mentioned that “where EdTech is positioned in that ecosystem kind
of depends on their specific role”. Many of the companies operating in this field enter in negotiations and agreements directly with schools. It is therefore important, Melanie Ehren argued, to “capacitate schools and teachers to think about what kind of agreements they need to enter into to ensure that these data are stored in ways that are beneficial [and complying with] children’s rights” and to “capacitate schools and teachers to ensure that data is used properly”. For Kruakae Pothong, there is a disconnection between the “actual purposes” a school wants a tool to serve and the purposes it serves in reality. Moreover, the push that schools feel to integrate technology, and the large number of options, often confuses schools. The shortage of independent evidence of technology benefits makes it difficult for schools to make fully informed decision and undermines opportunities for businesses to build trust with schools. That is why Kruakae Pothong advocated for a “standardised framework for assessing EdTech” which would be based on evidence and cover elements such as the purpose of use, respect for children’s rights, compliance with data protection and privacy, as well as accessibility. Barbara Wasson also underlined that “EdTech companies really oversell their products” and that, up until today, any “pedagogical benefits we know from [independent] research are very, very limited”. Tony Weir, Senior Inspector at the Irish Department of Education raised the importance of “hidden data” and the question around transparency. In his view it would be potentially beneficial to consider the creation of a European Centre, similar to the European Centre for Algorithmic Transparency (ECAT), that would be tasked to “prevent the spurious collection of data whereby data is just being collected because it’s useful to companies, doesn’t serve any educational purpose for the child or for the teacher or for the institution, but it’s good for these companies”. Related to this idea, Melanie Ehren mentioned that in the Netherlands, the Data Protection Authority is tasked with overseeing “the development and use of algorithms”. However, she added that ensuring transparency is very complex and that even these new actors “do not have the capacity to oversee every single use of data by EdTech companies”. That is why, Melanie Ehren said, it is important to empower those who use data in schools, through data literacy programmes, to be able to “understand some of these issues, ask those questions, be provided with tools to check the use of data”. On the other hand, and although agreeing with the need to upskill those using and collecting data in schools, Kruakae Pothong mentioned the limitations that the speed with which technology changes brings to training teachers and schools. That is why, Kruakae Pothong argued for the central role that education authorities should play in creating the structures to filter EdTech and setting the criteria for that. These criteria include the pedagogical value of the product, children’s rights, data protection, privacy, and data minimisation, all of which should be in place by design.

### Why is it important to ensure the exchange and interoperability of data in education systems?

Responding to the question about the importance of ensuring the exchange and interoperability of data, Barbara Wasson argued that in Norway students are given a secure single sign-on which corresponds to their unique identity and allows their anonymised data to be sharable. In this way, the Directorate for Education can conduct analysis and research on aggregated data coming from all students in the country. A similar process is also followed in the Netherlands. Melanie Ehren mentioned that the national bureau for statistics with the research council have developed a platform where data are combined to enable those with access to the data, to “see how students develop over their school career, look at wider patterns in the system over time and connect that to some other indicators around deprivation for example, or parental income, parental jobs, etcetera”. It is important to ensure the interoperability of education data as this allows authorities to “monitor indicators about quality, performance and other types of educational outcomes, and to understand differences between schools, between regions, to explain drop out, [and] early school leaving overtime, […] better projection and required capacity in the system and resources, number of student places required, [and] prediction of teacher shortages”. Moreover, Melanie Ehren adds that such capabilities can “be used to measure effects of interventions and new policies, both at the school level where schools, develop new programmes and want to know how effective these are to address students’ learning outcomes, but also at the national level, when we introduce new reform programmes to really understand the effects of those and how they play out, perhaps differently across different types of schools”.

“To have that kind of platform and infrastructure to connect these data in a very secure environment allows for analysis on trends in the education system beyond the pedagogical purposes of the classroom.” – Melanie Ehren
The experience in the UK is different, as Kruakae Pothong mentioned that there is no such national platform currently in the country.

What is often reported is individual schools “manually entering the data into the different systems and configuring their own data analysis to get these benefits of the insights from the data set”. However, as Kruakae Pothong argued, in this situation “the lack of interoperability means that there are a lot of missed opportunities and added administrative burden for schools and teachers [which] somehow goes against the several claims from EdTech providers that they ease the burden of vast administrative tasks”.

How can we design suitable governance processes for such an ecosystem?

Melanie Ehren argued that to design suitable governance processes, one should think of the various checks and balances that should be in place, as well as the regulatory incentives. These checks and balances should ensure that “the data is accurate, that rights are secured, but also, increasingly important with the use of algorithms, that diversity of data is properly safeguarded”. When designing such processes, she claimed, one need to not only consider the actors but also the parts of the process involved such as “collecting data, analysing data, comparing data, making it accessible to users and the decisions that we make on the data”. These represent different elements of the micro, meso, and macro levels mentioned earlier. Melanie Ehren conceded that thinking of the right checks and balances is a difficult task, but at least raising awareness of their need is important.

Adding to the elements that need to be checked, Kruakae Pothong, mentioned that is important to take into consideration the “impact of data processing” and the “business models of technology providers” as these can show the types of data being collected and the ways in which data is processed, shared, and used. She underlined that in most cases, it is the “hidden data” that is being processed and used from companies which are also those that schools and students are least aware of when it is happening. Kruakae Pothong also argued that some of these “hidden data” are arguably beyond the scope of data protection laws, making it, as mentioned by Melanie, difficult to check. Therefore, it is important to “ensure accountability throughout the product and service life cycle starting from the design and development stage”.

How do we best define the infrastructure to be put in place and what are the actors involved?

According to Barbara Wasson the infrastructure relates to the regulation, quality processes, as well as the competences of the people that participate in it. She sees infrastructure as part of the wider ecosystem discussed earlier. Moreover, Barbara Wasson mentioned that the infrastructure depends also on the education system of a particular country. For example, in Scandinavian countries like Norway, teacher autonomy is quite high and thus it is difficult to regulate certain things on a national level. However, she argued that areas such as data protection and impact analysis can be centralised in order to avoid overloading schools, municipalities, and regions which would be tasked to evaluate individually EdTech tools for their schools. Barbara Wasson added that it should not be the responsibility of the individual teacher to have the competence to properly evaluate such tools and technology. She argued that the infrastructure should maintain the right balance between autonomy of teachers and support from the state.

For Melanie Ehren the infrastructure is defined around the actors involved, the technology used, the oversight and safeguards in place, maintenance processes, as well as organisation and regulation elements to be considered. She also marked the importance of distinguishing between highly centralised and decentralised systems when defining the infrastructure as the two would look somehow different when in place. Assessing how “marketised” the system is, is another factor that Melanie Ehren considers in the infrastructure as this would show the place and influence of private actors such as EdTech companies.
She also shared that in the Netherlands, although the education system is very decentralised, the infrastructure around data use is centrally coordinated. This coordination is incentivised and financially supported by the government but parts of it are executed by “middle tier” organisations like the Councils for Primary and Secondary Education.

These organisations coordinate the work of the schools and allow for exchange and discussion around data comparability, use of platforms, etc. Melanie Ehren also mentioned that this coordination is possible due to the relatively small size of the country as well as the collaborative culture element that allows it.

Kruakae Pothong defined the infrastructure based on three elements. The first is “infrastructure for access” and refers to access to products, services, and data, the second is “regulatory infrastructure” which means “clear and comprehensive sets of rules and assessment criteria”, and the third is “infrastructure for safe and secure sharing”, making sure that data subject’s rights are respected across the interoperable data ecosystem. Kruakae Pothong claimed that when it comes to access, in the UK there are students that still do not have access to devices or the connection to broadband in schools is not stable. On the other hand, related to the third element of data sharing, Kruakae Pothong shared some initiatives that establish “data stewardships” which allow for safe exchange of data. A model of data stewardship acts like “a box” or a pool where all data is managed and shared on behalf of data subjects by a trusted intermediary for specific reasons (in the interests of data subjects and/or the public) with appropriate actors that request them. This could bring value to, for example, assessing connection and comparisons between data from the education and health sectors to benefit students. However, frictionless data sharing systems can pose risks related to commercial exploitation when not properly safeguarded and have not proven their value for data use.

Who should own the data and what are the actors to be involved in such process?

Addressing this question, Barbara Wasson underlined that it is important to clarify what the types of this data are. Although it is important when a school is signing a data agreement to consider information such as students’ responses to questions and final marks, it is even more important to consider other hidden types of data. Such data could include, for example, time stamps and the time a student takes to respond to a question.

Adding to that, Kruakae Pothong supported the idea that “students should own the data”. However, she argued that the concept of ownership is problematic when talking about minors as “ownership entails control” and that is hard in the context of education. Instead, Kruakae Pothong believes that proper literacy programmes should be in place and schools, teachers, parents, and children should engage in conversations “to enable children to exercise [some] control over their data”. Citing the Digital Futures Commission’s research, surveying 1014 students aged 7-16 in the UK, she noted that “50% of children did not get a chance to talk about any of the types of data, why schools are using different types of technologies in the school, what data are being collected [for] or which processes stand for what purposes”. Moreover, Kruakae Pothong shared that the same study found that “only 29% of the children surveyed said that their schools talked to them about why [the] school use technology for teaching and learning, and only 15% [of those] surveyed said that their school told them about their rights to opt out”. That, she argued raises the question of “if you don’t even know that you have the right to opt out, how could you exercise that right?”

Melanie Ehren argued that “we need to think about ownership in terms of different levels of data use and what that entails in terms of ownership”.

In her view, different levels carry different implications for ownership as well as purposes for use. Commenting to the discussion, Tony Weir stressed “how difficult it is now to retrofit a governance system onto processes that are already in place”. In his view, is “it’s not fair to allow the responsibility for this up to individual schools”. Tony Weir argued that schools do not have the technical capacity, the expertise, or the time for that as “they’re engaged in the really important business of teaching and learning”. He therefore agreed with Kruakae that clear procedures need to be in place from central education authorities that would be tasked to assess and filter the types of technology and specific tools allowed in classrooms. On the other hand, however, Tony Weir also emphasised the need to be careful about over-regulating in this area as this could pose the “risk that we will stifle innovation and stifle the interest of companies to become involved in this area”. In his mind, it is a very a delicate balance that need to be carefully kept and maintained by central education authorities and not left up to individual schools or teachers. Adding to this, Patricia Wastlau, Principal Advisor for Research and Innovation at European Schoolnet, brought up the issue of the growing amount of data collected and processed in education. In her view, “this could negatively impact students to be autonomous learners and teachers to be reflexive professionals”.

Meaningful and ethical use of data in schools.
Melanie Ehren agreed to that remark and added that it is important that we “acknowledge that [collected] data is what is easily measurable and that also skews sometimes our vision to what we find important in education, whereas much of it cannot be captured in data”. That means, she continued, that it is paramount that we ensure “that what we cannot measure is not left out of our thinking around quality and quality issues”.

Further reading suggested by the speakers:


Resources suggested by participants during the webinar for future reflection:

AVT - sluttrapport | Datatilsynet
Child Rights by Design - 5Rights Foundation & DFC (digitalfuturescommission.org.uk)
A Blueprint for Education Data (digitalfuturescommission.org.uk)
What do children think of EdTech or know of its data sharing? Read our survey findings – 5Rights | Digital Futures Commission