Introduction

Data-collecting technologies are increasingly ubiquitous in children’s lives, and their educational lives are no exception. The use and collection of information about children by technologies used to teach them, surveil them and make predictions about their abilities have profound implications for their privacy, well-being, safety and future prospects. The impact on children of data collection and use in the educational context has only become more concerning with the shift to remote learning compelled by the COVID-19 global pandemic. While remote learning programs have provided laudable benefits to some students when classes could not otherwise take place safely, the heavy reliance on educational technology (“ed tech”) prompted by the pandemic also raises a number of questions about how student data is collected and used.

A number of issues related to collection and use of student information merit focus from academia, civil society and policymakers. Relevant questions include:

- What kind of data is being collected from children via pedagogical technology, technology-based school infrastructure (such as scheduling apps, electronic hall passes, tools that process data about students in order to predict educational outcomes), and student surveillance programs? How is that data being used?
- What kind of evidence supports the pedagogical value of ed tech services?
- What are the regional variations to each of these questions, with an emphasis on ensuring that the risks and considerations for the Global South are not neglected, and the varying needs of countries are distinguished and addressed?
- Do any existing frameworks concerning the use of data in education sufficiently address key issues?

A range of evidence exists to answer these questions. This issue brief aims to provide a circumspect overview of that evidence, identify gaps in the research and unanswered questions and suggest possible steps and guiding principles that policymakers, educators, civil society and tech companies should consider.

Several key issues that are related but not directly connected to data governance in schools have already attracted considerable attention. These include the collateral effects of using ed tech on children’s well-being (such as online bullying, media literacy and attention problems); the idea of ‘digital citizenship’; and the implications of the digital divide for children in schools and homes with no or insufficient access to digital tools. On a global scale, privacy and other data governance issues, such as encoded bias and discrimination, have often taken a back seat to more seemingly fundamental questions of access to various technologies.
What kind of data is being collected from students, and how is it being used?

Ed tech has become an inextricable part of learning in many parts of the world, including technology that supports the ins and outs of school logistics, software programs that supplement or replace in-person teaching, the collection and analysis of student data to predict outcomes for the student or school, and the surveillance of students, ostensibly for their safety. Teachers are relying on software programs to provide students with lectures or assignments; some schools provide laptops, where monitoring of the student’s use of the device often includes their use at home. Schools are using technology to support or create basic infrastructure, like scheduling apps, bus pass apps, bathroom pass apps, and digital records systems. Other programs, like ClassDojo, can provide a disciplining function. As the COVID-19 pandemic abruptly interrupted in-person instruction, schools are relying on various remote learning, blended learning, or hybrid learning programs in order to continue educating their students from a safe distance.

To be sure, certain ed tech services can make logistical infrastructure and records management more efficient, enable schools to make their curricula more flexible, and offer other benefits. But many of these services aren’t as straightforward as a scheduling program or records storage system. The utility of these programs is highly variable, as is the evidence supporting touted efficiency, pedagogy, equity or safety benefits. Often, potential benefits will be accompanied by concomitant risks to students’ privacy, safety and their ability to be treated fairly, given the privacy and equity concerns attached to collecting information about children, examining their communications, tracking their movements and assessing them with questionable machine learning tools. Many of the ed tech programs described above can be co-opted to serve as surveillance tools, given the volume of data collected from students. Others are designed expressly for that purpose. In the United States, some schools have implemented social media monitoring programs that are designed to alert administrators to when a student might be a safety threat; thousands have deployed facial recognition programs that purport to be able to identify unauthorized intruders or weapons. Others have installed microphone-based ‘aggression detector’ systems that purport to alert authorities to threats ranging from school shooters to underage smokers. Certain schools in China have also deployed facial recognition systems on students, including to ‘monitor attention’ or isolate victims of abuse.

While these systems are generally implemented as a means to protect vulnerable students, they often put children at risk of being erroneously accused of misconduct and exacerbate the disproportionate policing of students of colour. Examples of technologies in use in the United States that risk subjecting children to racial discrimination include facial recognition programs with higher error rates for Black people and children, social media monitoring services and essay analysis software that struggle to correctly analyze African-American Vernacular English, and disproportionate rates of school discipline for Black children. These surveillance systems also endanger the privacy and free expression of all students without measurable safety improvements in return. The possibility that surveillance technology vendors will use the data they collect for other purposes beyond what is strictly necessary to provide educational service is also concerning.

The COVID-19 pandemic has prompted many schools to purchase additional surveillance systems with dubious value for public health and clear privacy and equity implications for the students being surveilled, such as facial recognition analysis to identify remote test-takers, eye movement tracking software to monitor students during remote testing, and thermal image scanners. The full range of these systems’ effects on student privacy, both currently and once children can safely return to classrooms, remains to be seen. But it seems likely that the use of new services, new data collection and use policies, and new vendor relationships will have at least some impact on what kinds of services schools are open to adopting, what students and parents are comfortable with and what educating young people will look like once in-person instruction is no longer constrained by COVID-19 safety concerns.
The pervasive collection of data on students invites the corollary inquiry of how that data may be used and abused, and by whom. In the United States, some universities have purchased information from data brokers and testing firms in order to evaluate incoming high school students, including to weed out students who will need financial aid. Research has similarly shown that data brokers — companies whose primary business model relies on selling, sharing and analyzing data about people, obtained through a range of means — are collecting data from high school students and selling lists of students based on actual or perceived characteristics to various companies, such as “students taking oral contraceptives” or “undocumented students”. Similar problems concerning data collected from students by private companies have been examined in South Africa, India, and elsewhere. This information about students, whether accurately or inaccurately inferred, could be used to discriminate against them, diminish their life choices, or even threaten their liberty. ‘Learning analytics’, or the automated analysis of data about students for the intended purpose of improving educational outcomes, has been embraced for a purported potential to provide better information as to which students need the most institutional support, and what kind of support they need. But many learning analytics tools, such as predictive systems that purport to analyze students’ skills, potential or academic work have the potential to perpetuate bias against students from certain demographic groups, as has been examined in other contexts in an expansive literature on fairness and machine learning. While this broader literature is highly relevant, the education-specific research is often lacking, particularly from a geographically comparative perspective. Automated assessments of students can wreak havoc on their education and well-being. When the International Baccalaureate (IB) was forced to cancel its annual placement exam globally due to the COVID-19 pandemic, it replaced the test with an algorithm that attempted to predict what students would have scored based on data points including “a student’s assignment scores, predicted grades, and historical IB results from their school”. More than 15,000 students signed a petition for an alternative assessment mechanism after many of them received significantly lower grades than anticipated. The methodology behind the IB assessment algorithm is not yet publicly known, but at least one of the known variables — historical IB results from the student’s school — invites arbitrary error and encoded bias.

For example, some companies offer services that purportedly identify students at risk of dropping out of high school, based on criteria described as “achievement, attendance, activity, demographic, and behavior”. But without a clear explanation of exactly what criteria are being used and how they are relied upon, it is difficult for a school to accurately assess the program’s efficacy or likely impact.

Information about students … could be used to discriminate against them, diminish their life choices or even threaten their liberty

Moreover, explicit reliance on race or socioeconomic status to assess a student’s ability to succeed academically is dehumanizing and would create a feedback loop that punishes students from certain demographic groups or who attend under-sourced schools. At the same time, other heuristics may unintentionally encode bias into the algorithm’s determinations, such as failing to account for the fact that ‘behaviour’ issues could be linked to an undiagnosed disability.

Another recent example in the United Kingdom demonstrates how biased or otherwise flawed algorithmic assessments can harm students. When the government’s Office of Qualifications and Examinations Regulation (Ofqual) replaced its in-person university entrance examinations with algorithmic predictions of how students would probably have performed, nearly 40 per cent of students...
received results that were lower than their teachers’ assessments of their qualifications, with students in expensive private schools receiving disproportionately higher predicted grades. The automated assessments relied in part on previous test results from the student’s school, which means that attending a school in a low-income area made it more likely for the student to receive a lower score. By the time Ofqual decided to allow students to receive whichever grade is higher between their algorithmic and teacher-predicted grades, many students had already lost offers of admission to university that were based on the algorithmic score, throwing their lives into disarray.

Optimism about the innovative potential of ed tech programs can also shield concerning problems of conflict of interest

The IB and Ofqual debacles demonstrate the potential dangers of incorporating algorithmic assessments into educational settings, particularly without rigorous safeguards. The use of new machine learning tools in education can exacerbate existing problems of structural bias and their use should be carefully scrutinized before adoption is even considered, including through algorithmic impact assessments, and if the program in question is adopted, ongoing audits.

What kind of evidence is available to support the pedagogical value and safety of various edtech services?

The concern of administrators, policymakers and researchers that students could be ‘left behind’ absent the incorporation of ed tech programs into curricula, is a ubiquitous theme in ed tech policy discussions. But the limited evidence as to the pedagogical efficacy of many of these tools and the rising evidence of the potential vulnerabilities they can create for learners requires a more cautious approach. Understanding the efficacy of ed tech interventions is vital, yet what evidence exists concerning the efficacy of ed tech programs on student outcomes is uneven. Certain approaches, such as maths programs used to complement in-person instruction, have shown modest benefits in learning outcomes, while online-only instruction has often been shown to have poor results for students. Remote learning programs prompted by COVID-19, for example, have had uneven results for students so far. Other programs tout pedagogical benefits that are unsupported by randomly-controlled trials and other rigorous evaluation methods, instead offering anecdotal testimonials and vague claims that avoid promising results the product does not deliver. Lack of adequate standards and safeguards can result in students being exposed to disturbing disinformation. Globally, the vast majority of ed tech research has tended to focus on digital literacy and internet safety initiatives in schools, rather than on evaluating the efficacy of ed tech being used to teach students, or the privacy and fairness implications of pervasive data collection and use.

How do various regions differ in their vulnerabilities and needs when it comes to student data governance?

Regional variations and socioeconomic disparities among student populations in different countries mean that observations about one set of students may not be wholly applicable to another. For students who lack access to fast and reliable internet entirely, concerns about the fairness implications of automated grading systems or the impact of social media
monitoring on their free speech may be less immediately relevant. School systems that lack the funds or local infrastructure to ensure students have access to the internet are less likely to turn to high-tech programs to surveil students through school-provided devices and accounts simply because those devices and accounts don’t exist.

For context, 43 per cent of the world’s households don’t have access to the internet, with lack of access disproportionately concentrated in developing nations. Only 20 per cent of people in the least-developed counties use the internet, as compared to 80 per cent of their counterparts in developed countries. A UNESCO policy review examining ICT readiness in education in Asian countries found a huge range in the countries surveyed; while Malaysia and Singapore reached 2.9 on a ‘policy readiness’ scale devised by researchers, Solomon Islands scored 0.1, and Palau 0.3, with a large spectrum in between. It also found that half of surveyed countries lack any security measures whatsoever for the technology used, and that all the countries surveyed lack adequate data on children’s behaviour and use of ed tech nationally and regionally. Similarly, a survey of teachers in the western region of Thailand found that those teachers are not being given enough resources to implement various ed tech programs within their curricula in a thoughtful way — nor did the teachers surveyed mention privacy or other data governance issues as a concern. A UNICEF report on the use of social media by children and adolescents in South Asia describes a “demonstrated lack of understanding” among parents and educators when it comes to privacy and privacy-protective measures. Device sharing is also more common in households in the Global South, which must be taken into consideration for policy recommendations for those students.

At the same time, income inequality, lack of support through social services and other structural factors mean that many children in wealthy countries don’t have the internet access that their schooling frequently requires. Even in highly-developed nations like the United States, millions of economically disadvantaged people, including children, lack access to affordable internet. About 58 per cent of American students report that they use the internet to complete their homework every day, while 15 per cent of households with school-age children do not have a high-speed internet connection at home, a percentage in which Black, Hispanic, and lower-income students are disproportionally represented. In a review of the role of learning analytics in African nations, Paul Prinsloo notes the gender disparity between internet access levels in South Africa, as well as a socioeconomic disparity. Regional differences are instructive, but countries are not monolithic, and avoiding inapt over-generalizations is key.

What policy frameworks exist to address these issues?

A number of researchers and policymakers have offered relevant suggestions for how to address the risks to privacy and fairness posed by ed tech services, though none comprehensively encompass all the relevant concerns or provide sufficiently detailed solutions. As noted above, the full range of children’s privacy concerns often takes a back seat to digital literacy and concerns about children’s safety online (such as harassment and vulnerability to physical harm from adults).

Schools, national and regional school boards and similar authorities need principles, guidelines and binding regulations

Much-needed guidelines that provide tips on helping children to use online services more safely are valuable, but structural data governance problems require structural, education-specific solutions, not solutions for individuals. A UNICEF research brief provides valuable insight into the regional disparities and lack of access to the infrastructure necessary for remote learning during the pandemic, such as internet, radio and television, but only briefly addresses privacy concerns. A 2018 Council of Europe set of guidelines on children’s rights...
in a digital environment addresses privacy, but still provides fairly high-level analysis. While previous frameworks have not addressed the full range of privacy and fairness issues in a comprehensive way, recent efforts provide reason for potential optimism. UNESCO has recently released a first draft of its guidelines on the ‘ethics’ of artificial intelligence, which includes a heavy focus on how machine learning can entrench and exacerbate existing biases and structural disadvantages for marginalized groups, while also addressing the privacy implications of machine learning technologies and discussing the potential risks of both problems for students. The Council of Europe is also working on new draft guidelines on children’s data protection issues in educational settings. These endeavours must continue and spread, as more comprehensive and detailed recommendations for the collection and use of children’s data in education that center privacy and equity concerns are sorely needed.

**Recommendations:**

**Better processes and more accountability are needed to govern the potential use of ed tech services.**

- Schools, national and regional school boards, and similar authorities need principles, guidance and binding regulations to ensure that any uses of technology in education address the inherent privacy, fairness, accountability, accessibility and equity issues that attach to the collection of student data. Those principles should be rights-based. Non-binding principles alone will be insufficient to correct current industry incentives to constantly track students and use their data in nefarious ways.

- Instead of framing the problem around whether students will be disadvantaged by a classroom that does not rely on ed tech services, policymakers, school administrators, teachers and parents should consider whether the service they’re evaluating has been demonstrated to achieve better learning outcomes for students, what new risks the service could introduce, and whether those risks are tolerable or capable of sufficient mitigation. They should carefully consider what specific problems would be solved by adopting a new ed tech service. And given the lack of evidence supporting the idea that increased reliance on technology improves student outcomes, educators and policymakers should be wary of assurances that schools will be ‘left behind’ by relying on in-person instruction rather than software substitutions for it. Instead, they should carefully consider the possible risks to student privacy, autonomy, free speech, well-being and safety alongside any potential benefits.

- Schools should consider using privacy and fairness impact assessments before adopting new technologies.

- Where applicable, public school systems should use public procurement criteria that include privacy objectives for ed tech contracts. Educational institutions need to be conscientious and transparent about the technology they require students to use.

- The role of ed tech services, when their use is deemed appropriate, should not be to ‘dehumanize learning’, but to augment rather than to replace analog methods.

**Additional focus is needed on the privacy and equity implications of ed tech services.**

- Student privacy protection policies must account for the fact that children are particularly vulnerable to their data being used against them later in life.

- High-level principles, voluntary best practices and legal standards should focus on constraining corporate behaviour, rather than privacy self-management by students, teachers and parents. The responsibility should not be on parents and teachers to protect children: data minimization, deletion, collection and use policies should be strict, and parents and teachers shouldn’t have to opt in to more protective settings.

- Researchers, policymakers and educators should not conflate media literacy and ‘digital citizenship’ initiatives with distinct concerns raised by...
collection and use of student data, such as privacy or discrimination against actual or perceived race, ethnicity, or nationality, disability status, gender identity, sexual orientation and other sensitive characteristics.

Reform efforts should be scrupulous about the appropriate roles of various stakeholders and the varying needs of different populations.

- Better norms of corporate responsibility regarding the collection and use of student data are necessary, but not sufficient. It may be valuable for educators, policymakers and civil society to consult with industry, but new research and frameworks should be driven by experts devoted to the public interest, rather than by the private sector.

- Companies developing ed tech products for children should consult with experts such as pediatricians, educational experts, child psychologists, children’s privacy experts and experts in children’s disabilities, as well as parents and children.71

- Student privacy protection policies need to balance protections for students and student autonomy, particularly in the case of older children.72

- Policymakers and school administrators should be carefully attuned to the different capabilities and needs of different school systems, including the likelihood of disparate privacy violations or disparate results from biased automated decision-making systems for children from historically marginalized groups.

- Researchers and policymakers should be conscientious about how regional differences create different needs and vulnerabilities, while being careful to avoid painting with too broad a brush.

More research is needed on a range of topics, including:

- The pedagogical value of using certain ed tech services in lieu of in-person instruction, when in-person instruction is safely available.

- The kinds of information collected from students around the globe, how it is used, the range of entities that are able to collect the information, and how they are able to do so;

- How the use of various surveillance technologies has spread in schools around the globe, such as social media monitoring and facial recognition technology (including affect analysis), and the effects of that surveillance on students, particularly students from marginalized groups;

- The use of automated decision-making in education and its impact on student privacy, dignity, safety and well-being;

- The effects of COVID-19-prompted remote learning programs on pedagogical outcomes, school and corporate privacy practices, and student, teacher, parent and administrator attitudes towards ed tech; and

- The variations across global regions and within individual countries of how schools are collecting children’s information and how they are using it, with a particular emphasis on practices in the Global South.73
Good Governance of Children’s Data project

The Office of Global Insight and Policy is bringing together 17 global experts in a project to explore trends in the governance of children’s data, including the tensions between different rules and norms, emerging concepts and practice, and implications for policy and regulation. Debate on the future of children's data affects a diverse range of issues, including data ownership and control, data fiduciaries, profiling for digital marketing purposes, child-friendly privacy notices, data erasure upon request, age verification, parental responsibility, data protection by design and default, algorithmic bias, and individual and group data.

The project aims to highlight the gap between the world we want for children and today’s reality, developing a manifesto on how children’s data could be optimally managed and what steps need to be taken. To help develop this manifesto, members of the working group will publish short analyses of different approaches to data governance.

Endnotes

1 For an example definition, see Hollow, D., Muyoya, C. and Brugha, M. (2016). Education Technology Map and Guidance Document, Jigsaw Consult 12 (December 2016), (“For the purpose of the mapping exercise the term ‘technology’ encompasses hardware, software, provision of connectivity, and specific assistive technologies for those with disabilities.”).


6 Ibid.


8 Fedders, at 1678-98; Gebhart, note 2.


18 Andrejevic and Selwyn, at 4-5.


25 Gasevic, D. (2018). Include Us All! Directions for Adoption of Learning Analytics in the Global South 4-5, in Learning Analytics for the Global South, Foundation for Information Technology Education and Development.


32 Ibid.


35 See, e.g., Gasevic et al, supra note 23.

36 Hoefnagl, C.J. (2016). Edtech: Promise and Peril 11-4, keynote address for the Technology, Law and Privacy Conference: Privacy and Education in a Social Environment (June 2016) Describes the lack of rigorous evidence to support claims made by edtech companies regarding their pedagogical efficiency.

37 Feddters, B. (2019).The Constant and Expanding Classroom: Surveillance in K-12 Public Schools, 97 N.C.L. Rev. 1673, 1701 (describing the evidence of edtech’s pedagogical value as “scant”, when it exists at all; Zeide, E. (2017). The Limits of Education Purpose Limitations, 71 U. Miami L. Rev. 494, 516 (noting the “few research studies showing that new technologies will provide better outcomes for students, schools, or the education system overall” and that “[n]ew any many data-driven education technologies have not been thoroughly vetted”); Hoefnagl, supra note 32, 11-14.


40 Hoofnagle, supra note 36, at 11-13.

41 Emerson, S., Giles, M., A Popular Online Learning Platform Was Actually Created by an Underground Religious ‘Cult,’ OneZero, (9 October 2020), https://onezero.medium.com/a-major-online-learning-platform-was-created-by-a-subterranean-religious-cult-whose-leader-has-ccc99e7adcaf-9a1f-8563-0b15-c963995907a2


44 See, e.g., Hoofnagle, supra note 36, at 8-11 (describing a Los Angeles School District initiative to provide students with iPads that went awry when, among other problems, the children received the devices before appropriate curricula had been developed to incorporate their use).


46 See, e.g., Koh, Y. (2020). New Mexico Sues Google Over Children’s Data Privacy, The Wall Street Journal (20 February 2020) (discussing ongoing litigation concerning claims that Google illegally tracked children through its G-Suite For Education software and school-provided Chromebooks);


49 Note 9, supra.


53 Ibid. at 6.


58 Axier and Anderson, note 56; see also, Goldstein, D., Popescu, A., Hannah-Jones, N. (2020). As School Moves Online, Many Students Stay Logged Out, The New York Times (8 April 2020), https://www.nytimes.com/2020/04/06/us/coronavirus-schools-attendance-absent.html. (“Titiayo Aluko, 18, a junior at Landmark High School in Manhattan, is one of the students trying hard to keep up with her classes who has been thwarted by her lack of access to technology. She has a district-issued laptop, but no Wi-Fi network in her Bronx apartment since her family had trouble paying the monthly bill. For classes like statistics and neuroscience, Ms. Aluko has tried to complete assignments and participate in video conferences using her cell phone, but that is sometimes impossible.”)

59 Prinsloo, Context Matters, at 29-30.

60 See, e.g., Prinsloo, Context Matters, at 25 (“It is impossible to provide an African perspective on the adoption of learning analytics, considering that the African continent comprises sovereign states, each with its unique regulatory framework, development agenda, information and communications technology (ICT) infrastructure, and state of adoption of online learning.”)


UNICEF works in the world’s toughest places to reach the most disadvantaged children and adolescents — and to protect the
rights of every child, everywhere. Across 190 countries and territories, we do whatever it takes to help children survive, thrive
and fulfill their potential, from early childhood through adolescence. And we never give up.

The Office of Global Insight and Policy serves as UNICEF’s internal think-tank, investigating issues with implications for children,
equipping the organization to more effectively shape the global discourse, and preparing it for the future by scanning the horizon
for frontier issues and ways of working. With dedicated expertise in seven policy areas — digital technology, human capital,
governance, the environment, society, markets and finance — the Global Insight team assists the organization in interpreting, and
engaging in, a rapidly changing world.

Office of Global Insight and Policy
United Nations Children’s Fund
3 United Nations Plaza, New York, NY, 10017, USA

© United Nations Children’s Fund (UNICEF), December 2020

The author thanks Linda Raftree, Jasmina Byrne, Monica Bulger, Steve Vosloo, Juan Pablo Ospino and Andaleeb Alam for their
helpful comments.

This is a working document. It has been prepared to facilitate the exchange of knowledge and to stimulate discussion. The text has
not been edited to official publication standards and UNICEF accepts no responsibility for errors. The statements in this publication
are the views of the author(s) and do not necessarily reflect the policies or the views of UNICEF. The designations in this publication
do not imply an opinion on legal status of any country or territory, or of its authorities, or the delimitation of frontiers.