RAND CORPORATION

Perspective

Expert insights on a timely policy issue

October 2018

Privacy and Interoperability Challenges Could Limit the Benefits of Education Technology

Katharina Ley Best, John F. Pane

The expansion of education technology is transforming the learning environment in classrooms, schools, school systems, online, and at home. Technology is playing an increasingly important role in education and has the potential to transform both the day-to-day experiences of students and the effectiveness of the entire system. The rise of education technology brings with it an increased opportunity for the collection and application of data. In this Perspective, we examine some of the possible implications of growing data collection and availability related to education technology. Specifically, we discuss potential data infrastructure challenges that could limit data usefulness, consider data privacy implications in an education technology context, and review privacy principles that may help educators and policymakers evaluate the changing education data privacy landscape in anticipation of potential future changes to regulations and best practices.

Education technology is a rapidly evolving, exciting area of technological innovation that promises personalized, targeted, interactive, effective, and even fun instruction for students with a wide range of skill levels and interests. Most stakeholders, including students, educators, families, software developers, and policymakers, agree that there are benefits to be gained from education technology, and there is no question that schools and teachers are increasingly adopting education technologies (U.S. Department of Education, undated). For students, education technology introduces a mechanism for instruction to be more tailored and individually engaging. Online or virtual education options could provide opportunities for students who do not traditionally have access to resources, facilities, or teaching staff for specialty subjects. For educators and parents, the move toward digital platforms and trackable metrics could revolutionize lesson planning, progress tracking, and educational cooperation and communication among the school, home, and other places of learning. For developers of educational

materials, the digital world opens up opportunities for new methods of engagement with students and teachers, while increased data collection and tracking can help developers improve and tailor their products. Finally, the introduction of education technologies and the information they collect could help researchers and policymakers develop a better understanding of what works in education.

Although education technology promises a data-driven, software-powered future for education, delivering on that promise requires a system for storing and safeguarding data in a way that benefits stakeholders at reasonable cost. Additionally, it requires a consensus about the future role of education technology data, limits on acceptable uses, and a shared understanding of the right to privacy of education technology users. In this Perspective, we examine two related challenges for a future driven by education technology and data: (1) the requirements for creation and storage of usable data, and (2) the means for safeguarding privacy in the face of increased data collection. In light of these challenges, we then discuss some principles and themes that may guide discussions around digital privacy, governance, and oversight to begin addressing these challenges.

The Challenging Education Technology Data Landscape

Much of the promise of education technology lies in its ability to use data to build a coherent picture of academic progress and the experiences that shape it. Such data can inform teachers, parents, and students about achievement and learning, and support better educational outcomes. It can also provide evidence to inform decisions about curriculum and technology adoption, and to feed into the continuous improvement of the technologies. Today's education technology environment does not yet support a highly coherent data environment. Stakeholders approach questions surrounding data collection, storage, and access with different and sometimes misaligned expectations, incentives, and goals. Additionally, decisions about technology procurement and use, data storage, data protections, and policies and practices surrounding education technology are often made at a very small scale; integration into a coherent local, regional, or national system may not be the top priority for local decisionmakers. These facts make it more difficult for all stakeholders to use education technology data to its fullest potential.

Different aspects of education technology data are useful or beneficial to students, families, developers, policymakers, researchers, and other stakeholders. These different data use cases naturally lead to different priorities for data collection and storage. Technology developers and vendors have the most direct control over what data are collected and stored, and may prioritize their own product-development goals, including building a functional and marketable software solution. Secondary goals may include satisfying educator demands for data-driven features or demonstrated efficacy for improving student outcomes. Currently, these goals lead to the creation of a data environment that is too low-level, too high-volume, and too difficult to use to support the needs of other stakeholders.

Education technology data are a fundamental part of the functionality of the technology solutions themselves. For example, data collected about performance on past problem sets may be used to shape the next individualized lesson, and information about mouse clicks in an online game may allow the software to respond correctly to user behavior. The type of data that help an individual education technology work is generally very low-level, machine-

readable information, such as mouse movements, keystrokes, and clicks. For many developers, especially early in the development of a product, it may be beneficial to keep as many of these low-level data as possible to aid in future development efforts, leading to high data volumes that further hamper usability for other purposes. Such data may also be unintuitive and difficult to aggregate into more-meaningful constructs. For example, a product may log the selection of a radio-button answer choice but fail to record sufficient information for another party to diagnose the student's misunderstanding that may have led to an incorrect answer. Developers may also fail to log important contextual information, such as school or demographic information, and even the ability to unambiguously link a student's data to external records for the same student.

One way to better align developers' data practices with the needs of other stakeholders would be to provide incentives to developers. More-educated, data-savvy consumers operating in a more-mature market may expect technology solutions to support dashboarding, tracking, effectiveness monitoring, and other research functions, linking new data practices to developers' sales.[1] The current system makes it difficult for educators to exert such market pressure as buying decisions, and requirements development often occurs at the local level. Individual vendors, school systems, and states are making isolated decisions about the future of education technology not only by selecting which technologies to adopt or discard but also by introducing practices, habits, regulations, and legislation that govern how education technology and related data are integrated with the educational experience and shared among stakeholders.

Designing a system that can support the many potential stakeholder uses for education technology data could be difficult and potentially expensive.

Designing a system that can support the many potential stakeholder uses for education technology data could be difficult and potentially expensive. The burden of data cleaning and translation of raw data into informative data elements is large, yet standardization is necessary to allow aggregation across different sources.[2]

Several groups have suggested frameworks for data standardization in education that could lead to improved utility and interoperability. For example, the School Interoperability Framework, developed by the nonprofit education data standards cooperative Access 4 Learning Community, offers an open specification for academic institutions, as well as a system for sharing data between institutions (Access 4 Learning Community, undated). The framework provides an extensible model into which schools can embed the data they want to store, track, and share. Similarly, the U.S. Department of Education–sponsored Common Education Data Standards (CEDS) initiative proposes a set of common data models that could help increase data interoperability among developers, data centers, researchers, and policymakers (Common Education Data Standards, undated). The CEDS initiative proposes standard data elements that stakeholders can choose to use within its system, as well as standardized data models that govern links between the various data elements. Ed-Fi, sponsored by the Michael and Susan Dell Foundation, has extended the CEDS

framework to develop the Ed-Fi data standard. Ed-Fi provides an education data standard that can be applied across education technology systems to provide secure data linkages and holistic data visibility (Ed-Fi Alliance, undated). The standard (and related data schema and application programming interface) are available free to those who join the Ed-Fi community, giving schools or school systems the tools necessary to create a coherent data infrastructure across their existing systems.

Education and outreach are a significant hurdle in the move toward a common data architecture. Such groups as Project Unicorn aim to address this issue (Project Unicorn, undated). Project Unicorn does not suggest a specific data standard, but it aims to educate stakeholders about interoperability, data security, and data privacy issues. It suggests a rubric for assessing adequacy of data standards and data protections, including privacy and data access controls. Another example effort is the Data Quality Campaign, which advocates improved data systems, data usage, and data safeguards, as well as related policies (Data Quality Campaign, undated). It emphasizes the importance of collecting meaningful data that can be used to support students, teachers, parents, schools, policymakers, and researchers. In doing so, the Data Quality Campaign has been exploring some of the privacy issues outlined in the next section. Groups such as these play an important role in helping the community develop a necessary balance between education technology's increasing reliance on student data and the protection of privacy.

The Challenges of Education Technology Data Privacy

The benefits of a common data architecture and data elements, such as those described in the previous section can only be realized if data can be shared effectively between organizations and individuals. Stakeholders must agree on not only what types of data are collected and stored, but also about how data should be centralized, aggregated, shared, or integrated with other information. Finally, stakeholders must agree on how and to what extent these data should be protected. All of these decisions require a joint understanding of expectations for data security and privacy. The data systems developed and populated by technology vendors, school systems, and researchers must be accessible to those who need them while also keeping private data safe from misuse. While better access to centralized education technology databases would help teachers, administrators, researchers, and policymakers, many of the changes that could improve data availability—increased and standardized data collection, simplified access controls, profile generation and tracking, online access—could threaten student privacy in both predictable and unpredictable ways. At the same time, increased centralization and standardization could help mitigate privacy concerns by more effectively distributing resources and best practices required for data security. While centralized data repositories make it easier to link data across sources and make connections that could violate privacy, centralization could also bring better safeguards, including sound access control policies, improved encryption, stronger privacy policies, and power to identify and address issues.

Schools have long held privileged access to student data for educational purposes, along with the responsibility to protect them. However, the growth of education technology is redefining the role of student data and what is required to protect the rights and privacy of the individuals involved.

Schools have long held privileged access to student data for educational purposes, along with the responsibility to protect them. However, the growth of education technology is redefining the role of student data and what is required to protect the rights and privacy of the individuals involved. There is already evidence that policies and procedures adapted from the days of paper records will not suffice as education moves online and students are accessing diverse data systems from around the world. For example, schools are extending or transferring data access rights outside their physical and electronic boundaries of control to make use of new and exciting education technology resources. The increased technical complexity of education technologies and associated data systems also mean that schools may no longer have the staff expertise necessary to act as data and privacy protectors, especially when they are members of small or underresourced local education agencies. Finally, future expectations about data and privacy protections might become even more sophisticated and demanding, further stressing the ability of school systems to fulfill their traditional roles as data collectors and protectors. With all of these changes, it may become unrealistic to expect schools to continue to bear the entire burden of protecting student privacy. Legislation; regulation; or local, state, and federal infrastructure could lead developers, governments, and other stakeholders to play a larger role.

While it seems unlikely that hackers would attempt to steal data on elementary students' usage of an online mathematics learning tool, we are learning that disclosure of even seemingly inconsequential digital data can have unexpected effects. The dangers of loosely guarded educational data may not be as obvious as they are for financial or medical data, but breaches could lead to unwanted profiling or solicitation. Education technology usage data could be combined with other information about an individual's computer or online session, such as IP address and login time. In extreme cases, these data could contain information on household dynamics, physical location patterns and habits, and other potentially sensitive information. As such, education data are intertwined in the larger societal debate regarding privacy in the digital age, discussed further the next section. Yet unique concerns arise from the fact that the data collected by education technology are academic in nature and capture information about individuals' abilities, skills, and behaviors. Traditional academic records, such as report cards and transcripts, capture some of this information and are generally releasable only with the permission of a guardian and/or student. Education technology data could provide a much greater level of detail, highlighting shortcomings or excellence in specific subject areas, capturing learning speed and even temporary achievement lags, and disciplinary records. Such information could be aggregated into a permanent, detailed profile that could

be used by colleges, employers, insurers, or financial institutions to judge individuals' future potential or risk, or for unwanted targeted advertising or solicitation. There is some evidence that student data collected by schools are already being solicited by and supplied to third parties interested in promoting their products (Herold, 2013).

Principles for Framing the Education Technology Data and Privacy Conversation

The debate about education technology data and privacy is part of a much larger struggle to find a balance between data usability and data-driven "progress" on the one hand and security and privacy on the other. From academic journals to popular media outlets, this question of how to balance data usability and privacy underlies one of the most lively contemporary policy discussions; the decisions we make about the meaning of digital privacy and the extent to which it must be protected may be one of our era's most impactful legacies (Welser et al., 2018). There is a vast, complex, and rapidly evolving literature on digital privacy.[3] Alan Westin of Columbia University defines *privacy* as "the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others" (Westin, 1967). Clearly, there is some agreement that education, and especially compulsory education at elementary and secondary levels, warrants some suspension of the complete right to privacy of students and parents. Students must turn in work to be graded, and grades and other outcome information is shared to support educational goals. Additionally, information about student progress and educational practices is a necessary foundation for the assessment and improvement of the education system for all students, making educational achievement data a sort of "public good" that cannot be kept completely private. Education technology mostly changes this calculus in terms of the extent of information that is collected; the introduction of new parties (developers) into the student-teacher-school relationship; and, importantly, the digital format of this information and the attendant ease of storage, replication and linking, and difficulty monitoring and controlling its use. A similar shift is occurring in many other domains.

The right to privacy is considered a basic human right by the United Nations (United Nations, 1948). Privacy is important not only because it allows individuals to control what information about them is shared, but because it also allows individuals to form trusting healthy relationships (Derlega and Chaikin, 1977); explore new ideas without fear of retribution; and pursue important financial, medical, or other interventions with reduced risk of negative consequences. Increasingly, preserving a right to privacy is an inconvenience and a barrier to full participation in a modern digital lifestyle, including social media, store loyalty cards, personal

The debate about education technology data and privacy is part of a much larger struggle to find a balance between data usability and data-driven "progress" on the one hand and security and privacy on the other. From academic journals to popular media outlets, this question of how to balance data usability and privacy underlies one of the most lively contemporary policy discussions.

electronic devices, digital utility meters, and credit cards—the list is growing and, in the future, might include education technology.

Many organizations and academics have compiled principles or tenets of privacy. Some of the most commonly cited include the Federal Trade Commission's (FTC's) five principles of Fair Information Practice and the Organization of Economic Cooperation and Development's (OECD's) eight privacy principles (Federal Trade Commission, 1998; Gerber, 2010). While the number and exact wording of these principles vary from organization to organization, the content is generally similar. Currently, these lists serve mostly as a reference and repository for best practices. In the future, they are likely to serve as the foundation for requirements that may actually bind education technologies to observing these principles. Currently, in education, rules governing human subjects research ensure that many of these principles are adhered to when education technology tools are used as part of research studies, but almost no similar requirements exist when data-collecting technology is used as part of the normal educational experience. This may change. Similar principles have recently been reflected in the European Union's General Data Protection Regulation (GDPR), which is likely to have a global impact on data privacy and increase requirements related to the adherence to these principles (European Union General Data Protection Regulation, undated). It includes stringent requirements for notification of data collection, establishment of consent, limitations on use and storage, and data integrity and security. This law requires that many of these principles be followed by *any* organization that collects *any type* of individual-level data for *any* reason.

In the United States, education data and school records are governed by the Family Educational Rights and Privacy Act (FERPA). The law gives parents or students who are 18 years of age and older some rights regarding access to school records, and it requires consent for disclosure of records to third parties. Several revisions to this legislation have been introduced but not enacted (for example, the Student Privacy Protection Act, the Student Digital Privacy and Parental Rights Act of 2015, and the SAFE KIDS Act). These update FERPA guidelines and, more explicitly, cover digital topics, such as online accounts or use in targeted advertising. Individual states have recently enacted more–rigorous education data privacy regulations, the most comprehensive of which may be California's Student Online Personal Information Protection Act (SOPIPA). SOPIPA transfers the responsibility for protecting student data directly to the companies and organizations that collect these data. It also prohibits specific uses of student data that are not related to education, such as advertisement targeting, creation of profiles, and selling of student data. This would preclude, for example, targeted advertising based on performance in an educational technology tool.

Rules governing education technology privacy could move in the direction of more-comprehensive laws, such as GDPR, in the future. Because so many data-collection organizations, especially digital ones, operate globally, the effects of these new European Union requirements are already being felt around the world and may soon be felt in the education technology space. In order to consider the implications of such requirements affecting the education technology space, we use the FTC principles to organize five general themes and discuss applicability of each theme to the education technology context.

Technology-enabled learning environments may look very different from what the parents and guardians receiving notices experienced as students. Tracking and comprehending the uses for and implications of several different technologies may become taxing.

Notice/Awareness

The Notice/Awareness principle states that consumers or users of a system that collects data about them must be made aware of the fact that data collection is occurring. This relates to OECD's "Collection Limitation Principle," which states that data should only be collected by lawful means with knowledge and consent. Meeting the Notice/Awareness requirement in the case of education technology is likely to be relatively easy as long as technology adoption is centralized and well documented. Students and parents can be made aware that education technology is part of the classroom or home learning environment; confirmation of this awareness, such as a parent signature, would be required. As the classroom technology environment becomes more complex, however, providing the kind of notice and awareness that may be required in the future could also become harder. As technology proliferates, schools and teachers may not be fully aware of all the potential data-collection mechanisms linked into their classrooms. Currently, teachers are free to adopt technologies on their own, technology use spans school-owned and personal devices, and more and more devices are digitally connected to the Internet of Things. These factors lead to a complicated technology landscape in which even the most immediate stakeholders do not fully understand the underlying architecture.

There is significant precedent for the need for notification in educational settings (signed report cards, permission slips, consent to participate in research, and others). Current FERPA regulations require notice only when data are *shared* with third parties, but new regulations in the vein of GDPR could require such notice (and consent) whenever data are *collected* about students. Technology-enabled learning environments may look very different from what the parents and guardians receiving notices experienced as students. Tracking and comprehending the uses for and implications of several different technologies may become taxing. Communications about technologies and their benefits and risks should be concise, easy to understand, and consolidated to cover in a coherent manner the comprehensive suite of technologies being used. These communications should be detailed enough to allow students and parents to understand potential risks associated with technology use, as well as to allow them to find more detailed information as they see fit.

Choice/Consent

This principle states that, in addition to simply being made aware of the fact that data collection may be occurring, users should be given a choice about whether or not they allow such data to be collected. Choice implies the ability to determine whether or not data are collected, while consent implies a requirement for a positive affirmation of the acceptance of data collection. Choice/Consent primarily relates to OECD's "Collection Limitation Principle,"

but touches on several of the other OECD principles (specifically, "Purpose Specification," "Data Quality," and "Use Limitation"). "Purpose Specification" states that users must be informed of the reason for any data collection. "Data Quality" implies that the data collected should be relevant to the stated purpose for which consent was given. "Use Limitation" requires that data be shared or made available only to support the stated purpose, with consent, and as allowed by law. In education technology today, a burden of Choice/Consent is already imposed on data collection for research purposes. However, if such rules as the GDPR come to govern education technology, adherence to this principle could be required even for everyday classroom use of data-collecting technologies.

The Choice/Consent principle may seem easy to implement. It is straightforward to send home a permission slip explaining that a new mathematics software is in use in the classroom. However, this principle may actually present a significant hurdle for several reasons. First, it is very difficult to determine whether, in light of all possible risks and future outcomes, *informed* consent has been granted. Choosing to opt out of a certain program or activity may also present significant additional burdens or risk. Opting out may not be a feasible or realistic option for students because it would limit their ability to participate in the class. Finally, opting out could be very disruptive to the education process if schools and teachers are faced with a wide spectrum of technology adoption and associated capabilities within a classroom. Given these constraints, it may be necessary to consider how schools might need a privileged role that, to some extent, can override the principle of Choice/Consent.

The constant development of new data analysis tools and methods further complicates adherence to the Choice/Consent principle. In order to truly give consent or be adequately educated about choices, students and parents must have a sufficient understanding of a very complicated and constantly evolving system. For example, new analytic methods could feature new uses for data that were not anticipated when consent was granted. One possible way to safeguard against changing data analysis capabilities is by introducing data expiration policies; a time limit on the storage of collected data could limit exposure to privacy concerns reliant on new data analysis techniques.

Specific to the education technology context, some additional barriers to adherence to the Choice/Consent principle may be the unique status of students and the fragmented nature of the education technology space. Because many students are minors, parents bear the responsibility for consent. Additional safeguards, such as requiring additional student assent, may be needed to give students themselves more control over their own privacy. Because many education technology developers are small companies and because use of multiple technologies may require consenting to many different privacy policies and data use agreements, it could be especially difficult for consumers to fully understand the ramifications of their choices and consent in this environment.

Because many students are minors, parents bear the responsibility for consent. Additional safeguards, such as requiring additional student assent, may be needed to give students themselves more control over their own privacy.

The principle of Choice/Consent poses one of the greatest challenges to developers, educators, and policymakers. In order to facilitate adherence to this policy, developers should ensure that privacy experts are consulted when designing new technologies, data storage systems, consent disclosures, and privacy policies. Educators and technology purchasers should hold vendors accountable for robust privacy practices and clear disclosure of data collection. Educators and policymakers should put pressure on education technology developers to improve standards in this area because this is often not a top priority today. Finally, educators and policymakers should work with developers, vendors, students, and parents to ensure that privacy policies and risks are successfully communicated in appropriate, comprehensible terms for the target audience.

Access/Participation

The FTC's third principle of Fair Information Practice is Access/Participation. This principle requires that those individuals whose information is being collected have some way of accessing, reviewing, and editing this information. This is related to OECD's principles of "Individual Participation" and "Openness," which state that individuals should be able to view and challenge data collected about them and be presented with a transparent picture of the policies and practices being used by organizations that house these data. The Access/Participation principle also touches on the concept of data ownership, implying that even if data are created or harvested by an organization or company, the individual about whom the data are collected reserves some rights to these data.

While adherence to this principle may seem relatively easy in theory, it could be difficult to establish a system that allows for such reviews. Building a secure system that allows stakeholders to review data collected about them without jeopardizing data security is a potentially complex and expensive undertaking. It may also be problematic to allow individuals to correct or edit their data records. Such alterations would need to be limited to correction of mistakes or misrepresentations, without undermining the effectiveness of education technology as a grading or scoring tool. Finally, developers may not want to disclose all data that are collected, for example, because intellectual property concerns or wariness of researcher scrutiny of the technology's effects on student outcomes.

To provide Access/Participation in an educational setting, schools need to guarantee the rights of students and parents or guardians to review and manage data. In order to pave the way toward satisfying the Access/Participation principle, school systems and policymakers must understand and manage data access and ownership agreements with technology vendors. Where data are housed and/or owned by a technology vendor, policymakers must ensure that agreements adhering to the Access/Participation principle are in place, that they are robust, and that there is a path

Information about student progress and educational practices is a necessary foundation for the assessment and improvement of the education system for all students, making educational achievement data a sort of "public good" that cannot be kept completely private.

for recourse in the event of a breach of these agreements. Adherence to the Access/Participation principle should require a user-friendly and accessible mechanism for reviewing and managing data, and information on how to access the system should be clearly communicated with parents or guardians and, when applicable, students. If data are housed or owned by the school or school system, schools and policymakers may retain more control over the system's ability to provide adequate Access/Participation, but the same requirements apply. Privacy experts should be consulted to determine how to provide the necessary access to student data.

Integrity/Security

Once stakeholders decide on the desired boundaries for data access in a given system, protections that enforce these boundaries must be put in place. The Integrity/Security principle states that data need to be accurate and secure—protected by access controls, encryption, and safe storage. The related OECD principle is called "Security Safeguards." This principle raises largely practical (but complex) concerns about how best to protect sensitive data, such as individually identifiable education technology records, from those who should not have access.

There are several common methods for achieving Integrity/Security. One is through anonymization or deidentification of data before they are stored or distributed, removing any information that could lead to the identification of a single user or small group of users. Anonymized data can be very useful in uncovering large trends or patterns. However, some research questions cannot be addressed with anonymized data. For example, anonymization can make it harder to link data across sources or integrate new information and observations over time. Additionally, it may be difficult to ensure that individual identities cannot be reconstructed with the addition of external information. Another common method for data protection relies on a clearance process to carefully vet which users should and should not be granted access. Unfortunately, such a process can be time-consuming, costly, and vulnerable to breaches, either intentional or not.[4] A third potential solution is to store data in a separate, protected environment accessible by only a limited number of individuals. Outside interested parties can submit code to be executed within this environment, receiving back output that aggregates across individuals. This method puts a large burden on a single organization to support all data inquiries. Stakeholders must agree to a funding and support mechanism for such an organization. Underpinning each of these methods for safeguarding information is a robust secure data storage infrastructure as well as encryption practices. The appropriate means for encrypting data and authenticating users are themselves significant and constantly evolving topics of discussion, and present a need to balance security and usability while protecting against faster and more-sophisticated means of attack.

There is precedent for the protected data environment approach within the U.S. education data community at the national level. The National Center for Education Statistics allows researchers to submit queries related to data on public and private schools, school systems, and colleges, but it is not clear how such a system could be expanded or replicated to include many diverse sources of and use cases for education technology data (National Center for Education Statistics, undated). This approach requires significant data centralization and standardization; as discussed earlier, this seems unlikely to be achieved in the education technology space in the near future. Yet a decentralized system where

Keeping personal data secure is difficult, even when all stakeholders have good intentions, qualified privacy and data security staff, and sufficient resources.

decisions about storage, protection, and access are made by individual developers or school systems is more likely to be vulnerable to attacks due to limited resources at the local level and does not afford the same data protections to everyone.

Along with Choice/Consent, the principle of Integrity/Security could be among the most challenging for the education technology community, and the increasingly digital world as a whole. Keeping personal data secure is difficult, even when all stakeholders have good intentions, qualified privacy and data security staff, and sufficient resources. Significant, high-profile breaches of data centers run by large, reputable organizations are common. In the past few years, the most notable examples are the Office of Personnel Management breach in 2015 and the Equifax breach of 2017 (U.S. Office of Personnel Management, undated; Federal Trade Commission, undated). Adherence to best practices in data encryption and data storage is critical, and extreme vigilance is required to ensure data security. Phishing attacks, human error, or use of default or common passwords are some of the many ways that adversaries could easily gain access to a system. Additionally, while current encryption practices are not particularly vulnerable to brute force attacks when used correctly, increasingly sophisticated algorithms and greater processing power could require new types of encryption practices in the future. Educators and policymakers should ensure that best practices are followed in data collection, storage, and transmission by anyone granted access to sensitive data. They should also require that school system employees are thoroughly and repeatedly trained on how to keep data safe. Qualified data security specialists should continually monitor and update the system. Finally, policymakers and educators should demand that guaranteed disclosure of any data breaches, however small, are part of vendor agreements.

Enforcement/Redress

The principle of "Enforcement/Redress" implies that individuals whose information is collected must have a means of filing a consequential complaint if their data are misused. This relates to the OECD principle of "Accountability," which states that data controllers must be held accountable for compliance and prevention of misuse. The possibility of strong repercussions for violations could help incentivize data holders to implement strong security systems. The idea of enforcement applies across the other principles, and implies that users should have recourse if any of those rights are violated. Implementation of the Enforcement/Redress principle almost certainly means introduction of legislation and regulation.

As with the Access/Participation principle, protecting the Enforcement/Redress principle may begin with understanding and managing data ownership. Policymakers and educators should ensure that data ownership and responsibility for data protection are clearly defined and well documented. They should also verify that data owners and administrators have included in user agreements and privacy policies the right to file complaints, as well as procedures for doing so. Finally, policymakers and educators

should continue to support legislation and regulation that protects the right to file an impactful complaint if data are misused.

The Future Digital World of Education

Discussions about data security and privacy are becoming more and more common in our increasingly digital world and will have an impact on the education technology space. The Federal Bureau of Investigation (2018) recently issued a public service announcement to increase public awareness of the risks to students. As consumers become more aware of privacy concerns, parents and students are likely to demand additional attention to these subjects. Indeed, the introduction of such sweeping legislation as the GDPR foretells that greater regulation of education technology data is inevitable. This Perspective discusses the various ways that general principles may need to be customized for the education context, to balance individual rights with the needs of schools to use data to support their educational missions. Standardization, integration, and usability will be important challenges for educators, policymakers, developers, and other stakeholders to grapple with as they attempt to reach consensus about feasible solutions that respect individual rights but do not unduly inhibit the potential for technology to transform education.

Schools and systems, in conjunction with parents and students, must develop clear expectations of privacy and data security practices. Developers must find ways to meet these expectations while also staying abreast of changes to policy and regulation. Mechanisms must be put into place to protect privacy and enable use of education technology data within the acceptable bounds. By thinking through the principles outlined in this Perspective, education technology stakeholders can begin building consensus on the right balance between privacy and the collection and use of student data. Understanding these principles and the possible future requirements of privacy protections can help educators, technology vendors, and other stakeholders more effectively communicate regarding the responsible future use of education technology.

Notes

[1] See, for example, survey results from the Bill and Melinda Gates Foundation's "Teachers Know Best" initiative (Bill and Melinda Gates Foundation, undated).

[2] Data cleaning is turning into a large and profitable industry taking up an estimated 60 percent of data scientist time (Redman, 2016). For more about technical challenges of data mining in education, see Bienkowski, Feng, and Means (2012).

[3] Key conferences in this field include the Symposium on Usable Privacy and Security (SOUPS), as well as the Network and Data Security Symposium (NDSS) and the related usable security (USEC) workshop.

[4] Privileged employees can inadvertently release private information, ss occurred when a staff member accidentally emailed Berkeley, California employees' Social Security numbers to the media (Neumann, 2013). For many examples of intentional insider misuse by privileged individuals, see Overfelt (2016).

References

Access 4 Learning Community, "The SIF Specifications," webpage, undated. As of August 14, 2018:
http://www.a4l.org/page/SIFSpecifications

Bienkowski, Marie, Mingyu Feng, and Barbara Means, *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics*, Washington, D.C.: U.S. Department of Education Office of Educational Technology, issue brief, October 2012. As of August 13, 2018:
https://tech.ed.gov/wp-content/uploads/2014/03/edm-la-brief.pdf

Bill and Melinda Gates Foundation, "Teachers Know Best," webpage, undated. As of August 13, 2018:
http://www.teachersknowbest.org

Common Education Data Standards, homepage, undated. As of August 14, 2018:
http://ceds.ed.gov

Data Quality Campaign, homepage, undated. As of August 23, 2018:
https://dataqualitycampaign.org/

Derlega, Valerian, and Alan Chaikin, "Privacy and Self-Disclosure in Social Relationships," *Journal of Social Issues*, Vol. 22, Issue 3, 1977, pp. 102–115.

Ed-Fi Alliance, homepage, undated. As of August 26, 2018:
https://www.ed-fi.org/

European Union General Data Protection Regulation, homepage, undated. As of August 16, 2018:
https://www.eugdpr.org/

Federal Bureau of Investigation, "Education Technologies: Data Collection and Unsecured Systems Could Pose Risks to Students," public service announcement, No. I-091318-PSA, September 13, 2018. As of September 17, 2018:
https://www.ic3.gov/media/2018/180913.aspx

Federal Trade Commission, "The Equifax Data Breach," webpage, undated. As of August 13, 2018:
https://www.ftc.gov/equifax-data-breach

———, *Privacy Online: A Report to Congress*, Washington, D.C., 1998. As of August 14, 2018:
https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-report-congress/priv-23a.pdf

Gerber, Ben, "OECD Privacy Principles," webpage, 2010. As of August 14, 2018:
http://oecdprivacy.org/

Herold, Benjamin, "'Impenetrable' World of Student Data Brokers a Major Concern, Study Says," *Education Week*, June 6, 2018. As of August 13, 2018:
http://blogs.edweek.org/edweek/DigitalEducation/2018/06/impenetrable_world_student_data_brokers_fordham.html?cmp=soc-edit-tw

National Center for Education Statistics, homepage, undated. As of August 16, 2018:
https://nces.ed.gov/

Neumann, Alyssa, "City Apologizes for Accidentally Sending out Social Security Numbers," *Daily Californian*, April 24, 2013. As of August 16, 2018:
http://www.dailycal.org/2013/04/24/city-apologizes-for-accidentally-sending-out-social-security-numbers/

Overfelt, Maggie, "World's Oldest Hacking Profession Doesn't Rely on Internet," CNBC.com, May 13, 2016. As of August 27, 2018:
https://www.cnbc.com/2016/05/13/a-surprising-source-of-hackers-and-costly-data-breaches.html

Project Unicorn, homepage, undated. As of August 13, 2018:
https://www.projunicorn.org

Redman, Thomas, "Bad Data Costs the U.S. $3 Trillion Per Year," *Harvard Business Review*, September 22, 2016. As of August 14, 2018:
https://hbr.org/2016/09/bad-data-costs-the-u-s-3-trillion-per-year

United Nations, Universal Declaration of Human Rights, Article 12, December 10, 1948.

U.S. Department of Education, "Use of Technology in Teaching and Learning," webpage, undated. As of August 13, 2018: https://www.ed.gov/oii-news/use-technology-teaching-and-learning

U.S. Office of Personnel Management, "Cybersecurity Incidents: What Happened," webpage, undated. As of August 13, 2018: https://www.opm.gov/cybersecurity/cybersecurity-incidents/

Welser IV, William, Rebecca Balebako, Cameron Colquhoun, Osonde Osoba, "The Latent Potential of Privacy Technologies: How Our Future Will Be Shaped by Today's Privacy Decisions," in Zachary S. Davis and Michael Nacht, eds., *Strategic Latency: Red, White and Blue: Managing the National and International Security Consequences of Disruptive Technologies*, Livermore, Calif.: Lawrence Livermore National Lab, February 2018, pp. 171–187.

Westin, Alan, *Privacy and Freedom*, New York: Antheneum Press, 1967.

About This Perspective

The expansion of education technology is transforming the learning environment in classrooms, schools, school systems, online, and at home. The rise of education technology brings with it an increased opportunity for the collection and application of data, which are valuable resources for educators, schools, policymakers, researchers, and software developers. RAND researchers examine some of the possible implications of growing data collection and availability related to education technology. Specifically, researchers discuss potential data infrastructure challenges that could limit data usefulness, consider data privacy implications in an education technology context, and review privacy principles that could help educators and policymakers evaluate the changing education data privacy landscape in anticipation of potential future changes to regulations and best practices. This research was conducted in RAND Education.

RAND Ventures

The RAND Corporation is a research organization that develops solutions to public policy challenges to help make communities throughout the world safer and more secure, healthier and more prosperous. RAND is nonprofit, nonpartisan, and committed to the public interest.

RAND Ventures is a vehicle for investing in policy solutions. Philanthropic contributions support our ability to take the long view, tackle tough and often-controversial topics, and share our findings in innovative and compelling ways. RAND's research findings and recommendations are based on data and evidence, and therefore do not necessarily reflect the policy preferences or interests of its clients, donors, or supporters.

Funding for this venture was provided by gifts from RAND supporters and income from operations.

Acknowledgments

The authors are grateful for the insightful comments and suggestions provided by Cathy Stasz, John Davis, and Jennifer Bell-Ellwanger. The authors welcome reader comments and queries.

About the Authors

Katharina Ley Best is a full operations researcher at the RAND Corporation. Her research interests focus on applications of operations research and financial engineering methods. Best's recent work has focused on Army strategic planning and decisionmaking under uncertainty, modeling of military career paths and grade structure, and education programs for military spouses.

John Pane is a senior scientist who studies technology innovations in education using rigorous research methods. His recent work includes the first large-scale evaluation of schoolwide personalized learning and several efficacy studies of intelligent tutoring systems. He held RAND's Distinguished Chair in Education Innovation from 2015 to 2018.

For more information on this publication, visit www.rand.org/t/pe313.

www.rand.org

PE-313-RC (2018)