Workshop on Big Data in Education

Balancing the Benefits of Educational Research and Student Privacy

Lessons Learned from Education Stakeholders: Panel Summary

Bill Fitzgerald, *Common Sense Media*

NATIONAL ACADEMY of EDUCATION

Lessons Learned from Education Stakeholders
Panel Summary

Bill Fitzgerald, Common Sense Media

Workshop on Big Data in Education:
Balancing the Benefits of Educational Research and Student Privacy

National Academy of Education
Washington, DC

NATIONAL ACADEMY OF EDUCATION 500 Fifth Street, NW Washington, DC 20001

Additional copies of this publication are available from the National Academy of Education, 500 Fifth Street, NW, Washington, DC, 20001; https://www.naeducation.org/bigdata.

Printed in the United States of America

Suggested citation: Fitzgerald, B. (2017). *Lessons Learned from Education Stakeholders: Panel Summary*. Washington, DC: National Academy of Education.

Lessons Learned from Education Stakeholders: Panel Summary
Bill Fitzgerald, Common Sense Media

Panel:
Bill Fitzgerald, Common Sense Media (Panel Chair)
Chris Dede, Harvard Graduate School of Education (Panel Moderator)
Olga Garcia-Kaplan, FERPA | Sherpa
Jim Siegl, Fairfax County Public Schools (Virginia)
Amelia Vance, National Association of State Boards of Education

INTRODUCTION

This panel focused on articulating the tensions between the potential for research using big data and the privacy of people whose information make up data sets. The discussions surfaced several major themes; while these themes were not explicitly used as points of organization during the session, they help structure the major ideas that arose in dialogue. For each of these themes, a variety of opinions existed within the room, and in this summary differing opinions are included to represent contrasting viewpoints. Toward that end, this summary should be understood as a reporting of ideas that were covered, not as the group's suggestions or areas where consensus was reached. This session was an environmental scan rather than a decision-making process, and the summary attempts to reflect this.

Researchers are committed to doing the right thing, but the path toward the "right" thing is not always as obvious as we would like to believe. Within the research community, we recognize that most academic research is conducted for the greater good, and that researchers doing the work have ethically sound motives. However, this needs to be communicated clearly to those who have supplied data to avoid even the impression of a "conspiracy of good intentions."

POTENTIAL OF BIG DATA

The power and potential usefulness of big data are clear, and this session did not dwell on the myriad uses of big data in research. As discussed in the opening remarks of the session, the value of big data research includes insights gleaned from large data sets, but also the use of big data to identify questions that could be explored in more detail through other research studies. We are in the early stages of determining how to use and analyze big data to improve education. Dr. Chris Dede analogized educational big data to the telescope and the microscope. These tools did not create new things for us to see, but they allowed us to see a greater depth and breadth of things that had always existed. The scientific community, however, still requires time to figure out how to use the data uncovered by these instruments and analyze the newly visible data. Data-intensive research has the same possibilities to enhance educational outcomes; however, as Dede pointed out, stars and bacteria do not have the same privacy issues as student data.

The vision for big data is that it collects a lot of information about student learning that can be reported to students, parents, and educators to assist with decision making. As we use big data for different types of research, we have an obligation to use the findings we develop to address and reduce traditional educational gaps. Consequently, addressing issues of equity should be a major goal of big data research.

PRIVACY AND SECURITY CONCERNS

When thinking about data privacy, we often lose sight of the fact that one element of privacy involves controlling access to the story of a specific person. As a result, talking about privacy without including the need to protect learner agency, or the ability of the learner to control what is collected, what is stored, and what is shared, is an incomplete conversation.

Additionally, I have heard researchers—in different events spanning several years—making the statement that their data are not "interesting" enough to merit the time and effort to access it. These attitudes toward data security ignore the fact that motivations and mechanics

of unauthorized access vary widely, and the perceived "interest" of a data set should not be considered a form of protection. For researchers wanting trust, they need to be aware of and in line with current security and privacy practices. This includes knowing how to avoid security risks and embracing setups that use defense in depth.

Corporate data collection, such as in the consumer space—especially what is done by the gaming industry—goes far beyond what is collected in educational research. The broad range of data collected by video games and the opportunistic nature of that collection, however, does not alter the reality that choosing to play video games is different than compulsory public education, and that opportunistic privacy practice in the corporate sector does not alter the ethical obligations of academic researchers in education. The pendulum of privacy awareness seems to be swinging back toward greater concern for basic privacy protections.

Institutional review board (IRB) review often lacks people with the technical or professional backgrounds to assess whether or not a research project has adequate technical or procedural safeguards in place. One of the more visible examples of this was the Facebook mood study, where for a week Facebook presented differing ads to Facebook users (some with more positive or negative comments), and the subsequent study examined if that in turn impacted the language or mood choices of the users' posts. One author of the paper was a university professor, and the university's IRB deemed that review was not necessary because its professor was not directly involved in collecting the data. This approach to determining whether or not IRB review is even required could theoretically stretch to cover many partnerships with industry. As was seen with the fallout from the Facebook mood study, the rationale of the IRB did not satisfy skeptics. The gray area of IRB oversight is exacerbated by for-profit IRB providers that seem to market themselves based on a fast turnaround rather than a thorough review. In short, IRB review can be applied inconsistently across academic research.

Siegl raised the idea of "reasonable privacy" as a counterpart to the ever-evolving concept of "reasonable security." The idea of reasonable privacy can be applied to protecting the identity of people within deidentified and aggregate data sets, but the concept of reasonable privacy can only evolve if people within academia articulate and adhere to consistent standards about what is considered "reasonable." Time and technology have shifted some key definitions used in privacy law and thus the notion of reasonable privacy. As noted by Siegl, in the 1970s when the Family Educational Rights and Privacy Act (FERPA) was written, "deidentified" meant "not recognizable by someone with knowledge in the school community." With big data, the Internet, and cheap computing power, we have redefined the notion of "school community" and "knowledge of the school community." We also have increased the amount of knowledge easily accessible for analysis.

POLICY AND COMPLIANCE LANDSCAPE

According to Vance, since 2014, 19 states have passed laws that contain at least a single provision targeted at research. These new laws—and the myriad other laws that have been discussed but not enacted—are indicators of the existing fear and mistrust about the role of the government in education and research. Additionally, 36 states have passed 85 new laws addressing student privacy since 2013. While some of the laws that have passed recently have fairly reasonable requirements around governance structures for research, other laws address the structure of data that can be made available to researchers. For instance, a 2014 Kansas law requires parental permission for any study or, alternatively, that all data be aggregated

before they are made available to researchers. Local educational agencies and state educational agencies, with their limited resources, do not have time to aggregate data in a useful form for researchers, significantly limiting research in Kansas. And some of these laws have unintended consequences, such as banning necessary data collection for teacher certifications or banning the necessary documentation and sharing of information concerning students with disabilities, which must be later addressed and fixed by legislatures.

At the federal level, bills have been introduced or are in earlier stages that rewrite or enhance FERPA. For instance, federal legislative efforts include provisions that forbid research unless there is a direct benefit to the subject of the research, require opt-in for all research, require deletion of student records when the student leaves an institution, and prohibit federal funding for research that uses psychological data.

Another potential collision course on the horizon, pointed out by Vance, involves the social/emotional learning (SEL) standards being rolled out in many states. These standards are supported by the Every Student Succeeds Act. A consortium of 8 states are currently collaborating on standards, and an additional 11 will have access to the consortium's work. The implementation of these standards will be accompanied by data collection to help understand how they are being implemented, and whether certain implementations are more effective than others. The sensitivity of children's social and emotional data has the potential to increase existing privacy and security concerns, and there are already growing objections to any collection of SEL data from children.

While many academic researchers have expressed confidence that policy makers understand the value of research, that assumption is not safe to make in the current political climate. For example, at the federal level, Lamar Smith, the chair of the House Committee on Science, Space, and Technology, has had ongoing clashes with and demonstrated disrespect to the scientific community, and the current head of the Environmental Protection Agency has stated his skepticism about climate science. These examples demonstrate how some policy makers perceive and vilify researchers.

State and local educational agencies have been playing catch-up over the last several years to create and improve data collection policies. Prior to that, there was a lack of transparency in how they collected and shared information for research purposes. Now, however, many school districts (in general and anecdotally) have policies in place guiding working with researchers and IRB processes. Many state agencies, however, still do not have a formal process governing data collection or use in research.

Furthermore, many state and local educational agencies use compliance with FERPA as their guiding rule, and do not do much beyond that. Other federal requirements—such as the Children's Online Privacy Protection Act (COPPA) or Protection of Pupil Rights Amendment (PPRA)—are not widely known or understood, and states have a range of additional requirements. Moreover, many terms within these regulations are defined differently by different school districts. For example, under FERPA, who is determined to be a "school official" varies from district to district, as does who can contractually bind a district. While these definitional differences may not be directly relevant to research concerns, they highlight an often-overlooked element of the current landscape: districts have broad autonomy to create rules and often lack the support staff to use that autonomy effectively across different situations.

COMMUNICATION NEEDS

As Garcia-Kaplan noted, "big data can tell big stories and those big stories matter." However, the follow-up question is, "How do these stories matter to kids?" It is critical that researchers share the goals and benefits of their research with the sources and stakeholders of the data they drew upon.

While the benefit of research is clear to researchers, it is often not made clear to or shared with the subjects who gave their time, their stories, and their personal information to the research. This is often compounded when data are shared with other researchers, and the subjects are not notified. The tensions in public K-12 education are magnified because student attendance is compulsory but individual data are private to a student. Even though attendance is mandatory, participation in research is often seen as something where learners and their families deserve a choice.

How does research change—and how does our thinking about data collection and privacy change—if we explicitly ask what people participating in research studies need, and how the researcher can provide it? When we have the frame of how kids and parents can benefit, the issue of privacy shifts. It is no longer a matter of accessing data; it is a matter of sharing information and insights to help people learn.

From a communication perspective, many people are okay with data sharing, but they want to feel a level of involvement in the process. Research subjects will likely be open to a range of options regarding their data in studies, but they also want to know that their voice and consent matters.

In cases where academic researchers partner with industry, how do academic researchers differentiate themselves from corporate researchers? Many vendors include provisions in their terms of service that explicitly claim the rights to use data for research. Some of these same vendors also commit to strict data sunsets, which many academic researchers do not even consider. Increased transparency on the part of academia can help answer some of these questions, but the additional transparency will almost certainly bring additional questions. If we are to take a lesson from other industries and the patterns we see from legislative activity, academia would be well advised to proactively embrace additional transparency and examine their data handling and privacy practice.

In some communities, considerable resistance to academic research already exists. Researchers have an obligation to assume good faith and rational agency of people who disagree with data collection and use in academic research. We can disagree fully with the arguments of people who oppose academic research and simultaneously work to understand the nature of these objections. Research conducted in public schools, on children who are required to be there, potentially carries a different set of obligations and concerns than research conducted in other contexts.

Vance posed three issues that have been repeated frequently by parents and policy makers that academic researchers should be prepared to address:

- What is education research, and why should I care about it?
- Researchers are able to get access to student data and use it for whatever they want.
- Parents should always be allowed to opt their child out of research that will not directly improve their child's education or help their child in some way.

These questions and concerns encapsulate many of the fears and critiques of research. They touch on privacy concerns, the relevance of research, learner agency, the question of who should benefit from research, and parental rights. These questions or concerns might feel naive within academia; however, outside academia, many of the objections and critiques of educational research build on these three core concerns.

During the question-and-answer period of the panel, a participant observed that many concerns about privacy in big data research are also about "ethics and decision making and who has control over what happens to students, especially the information being fed back for use." As academic researchers talk with more people outside academia about research, many of the conversations will address who has the right to say what matters, and who has the right to access stories that are not theirs. While there are security aspects and transparency aspects to the conversation, we cannot overlook the human concerns in which some of the more technical or arcane conversations are grounded.

NEXT STEPS

While what follows is certainly not exhaustive, it describes some potential next steps to ensure that necessary research occurs while protecting privacy and security. Two frameworks were introduced by Siegl and Vance to illustrate how to explain data use and how to anticipate and safeguard against concerns.

Siegl discussed a framework using the acronym CUPS:

- **Collection**—what is collected, by whom, and from whom?
- **Use**—why are the data collected, and how will the data be used? Will participants be informed about use?
- **Protection**—how is access to the data restricted? These protections include both technical and human concerns; and
- **Sharing**—how can data be shared, and with whom? Are participants in the original study informed if/when data are shared?

If researchers describe to participants and other interested parties how data will be collected, used, protected, and shared, the intentional transparency will help allay many concerns about potential misuse. In addition to being a useful communication tool, defining these elements is also good practice. The exercise of explicitly defining these elements also will help identify potential issues that could arise if or when privacy and security protections break down.

The second framework, introduced by Vance and attributed to Siegl, helps categorize how privacy and/or security protections break down. Generally, issues arise when we see

- Behavior that breaks existing law;
- Behavior that violates security and privacy best practice, but that is within the law; and
- Behavior that breaks trust and expectations.

Issues can fall into more than one of these categories, but they are all serious, and even legal behavior that breaks trust or expectations can have serious repercussions. If researchers are intentional in how they define CUPS, they will often be able to flag potentially problematic behaviors before they become an issue.

Other ideas discussed to minimize potential privacy issues are included below. These ideas all would require additional thought and planning before being implemented.

- **Create centrally accessible stores of administratively collected data that have been deidentified.** Determining what fields should be included, how data should be deidentified, who should administer the data stores, and sources of the data would require collaboration between multiple institutions. In addition to security and privacy concerns, a large accessible data store containing student data could also pose communication risks related to how people trust or mistrust a single entity amassing a large data set on students.
- **Schools and districts could be more proactive around publishing research and studies that they participate in,** complete with IRB information, data collected for the study, and the goals of the study. This is comparable to how schools have improved transparency in publishing the list of vendors who receive data under the "school official" exemption of FERPA.
- **Academic researchers and institutions must be more proactive in their security and privacy practice.** An example discussed during the workshop was the number of researchers who connected to a public wireless connection in the workshop room that had no password, without using a VPN to protect their communications.
- **Laboratories that collect and store sensitive information must obtain regular security audits.** These audits should be accompanied by professional development for staff in data handling.

CLOSING THOUGHTS

Discussions around privacy, security, the role of IRB review, and how to communicate the value of educational research to people outside academia who might not value or appreciate the role of research flowed through several of the panel discussions. There is no monolithic viewpoint that will fully address the multiple issues discussed during the National Academy of Education workshop. Within the room, there appeared to be a strong desire to continue to examine issues related to privacy, and to ensure that educational research sets a high standard for respecting the privacy of people who participate in research. Additional conversations on privacy and security would be useful to identify areas where consensus can be reached and to identify some best practices that could be shared among researchers.