Statement of Paul Ohm
Professor, Georgetown University Law Center
Member, Commission on Evidence-Based Policymaking

Testimony for the Hearing on
Protecting Privacy, Promoting Policy:
Evidence-Based Policymaking and the Future of Education

Before the
Committee on Education and the Workforce
U.S. House of Representatives
January 30, 2018

Chairwoman Foxx, Ranking Member Scott, and Members of the Committee, I appreciate the opportunity to be here with you today to discuss the relationship between evidence-based policymaking and privacy protection in the education context.

I am a Professor of Law at the Georgetown University Law Center. My scholarship and teaching focus on information privacy, computer crime, and technology and the law. I have held positions relating to privacy law and policy in the federal government three times. Most recently and of most immediate relevance, from 2016 to 2017, I served as a member of the Commission on Evidence-Based Policymaking, which issued its final report on September 6, 2017.[1] In addition, from 2012 to 2013, I served for ten months as the Senior Policy Advisor for Privacy for the Federal Trade Commission; and from 2001 to 2005, I was a trial attorney in the U.S. Department of Justice's Computer Crime and Intellectual Property Section. I make these comments in my personal and academic capacity and they do not necessarily represent the views of any of the organizations listed above.

In these brief remarks, I would like to give some context for the Commission and its work; propose four privacy principles that should sit at the heart of evidence-based policymaking; demonstrate how those principles helped shape the Commission's report; and explain how these insights fit into education policy in particular.

1 Congress Placed Privacy at the Center of the Commission on Evidence-Based Policymaking

As you know, the Commission on Evidence-Based Policymaking ("Commission") was created by a statute initially co-sponsored by Speaker Ryan and

[1] The Promise of Evidence-Based Policymaking: Report of the Commission on Evidence Based Policymaking (2017) (*hereinafter* Commission Report), *available at* https://cep.gov/content/dam/cep/report/cep-final-report.pdf.

Senator Murray and signed into law on March 30, 2016.[2] The statute allocated the Commission only eighteen months to produce its final report before it expired, and it was through some truly heroic hard work of the Commission's staff that we met the deadline and produced the report, which is available at https://cep.gov.

The statute charged the Commission to "conduct a comprehensive study of" the data activities of the Federal government; "determine the optimal arrangement for which administrative data" of the Federal government "may be integrated and made available to facilitate program evaluation, continuous improvement, policy-relevant research, and cost-benefit analyses by qualified researchers and institutions"; and make recommendations toward these ends.[3]

In addition, Congress's charge placed privacy at its center. First, the charge required the Commission to "weigh[] how integration might lead to the intentional or unintentional access, breach, or release of personally-identifiable information."[4] Second, the statute mandated that one-third (five of fifteen) of the members of the commission "shall be . . . expert in protecting personally identifiable information and data minimization."[5] My appointment designated me as one of these five experts.

I applaud Congress for making these choices, as the work of the Commission was steered toward taking privacy seriously by them. Our Commission discussed privacy at every single open and closed meeting, and our final report bears the fingerprints of Congress's institutional design choice by placing privacy at its core. This validates the theory that institutional design matters, and I encourage the members of this Committee to find other opportunities to "bake privacy in" to other entities you create and oversee, including government agencies.

In my time on the Commission, I considered it my statutorily backed mission to educate my non-privacy-designated colleagues about the theory, practice, and technology of privacy. To be candid, I became like a nettlesome thorn in the sides of some members of the commission, many of who are prominent and talented social scientists who had spent their careers engaging with privacy concerns only at some distance. I am grateful that my colleagues looked past their personal irritation with my persistent single-mindedness to take seriously my concerns, and although the final report is by no means perfect, I was happy to place my signature upon it.

2 Four Principles to Respond to the Unavoidable Tradeoffs Between Social Science Research and Privacy

At bottom, I think of the phrase "evidence-based policymaking" as mostly an exercise in political branding, a euphonious label for a simple, almost obvious, set of ideas: we ought to measure what government does with rigor and integrity, and we

[2] Evidence-Based Policymaking Commission Act of 2016, Pub. L. 114-140 § 4(a) (2016).

[3] *Id.*

[4] *Id.* § 3(a).

[5] *Id.*

ought to adjust our policies when these measurements reveal that our plans are not working. Stated at this level of generality, who could possibly oppose it?

The problem is that when we move past this level of generality, the situation becomes far less clear and far more contestable. There are proposals that might fly under the broad banner of evidence-based policymaking that would not do enough to protect privacy. I come to this conclusion from my research into the vexing tradeoffs between data analysis and privacy protection.[6] The hard work required of evidence-based policymaking requires us to grant researchers access to data about the private activities of the citizens and residents of this country. These data can sometimes—not always—reveal the sensitive activities, habits, relationships, and even thoughts and aspirations of individuals, the kind of data that can lead to great harm if allowed to disseminate too far.[7]

In education, we of course are often talking about data about children, our most precious and vulnerable population. Children deserve a heightened level of privacy protection, so we must take special care to scrutinize new research studies that place their privacy at risk.

My work suggests four lessons about the privacy impact of data analysis that I hope this Committee keeps in mind as it considers the tensions between evidence-based policymaking and privacy protection: the tradeoffs are unavoidable; there are no silver bullets; privacy requires friction; and privacy requires structural constraint, not merely rules of the road. Let me elaborate on each of these in turn.

If there is one consistent theme in my research, it is the proposition that data analysis and privacy invasion are two sides of the same coin.[8] The very same mechanisms that let researchers use data to learn useful information about people and programs can also, in the wrong hands, lead to serious and harmful invasions of privacy. Often, the only way to distinguish between the two is to examine the subjective intent of the person looking at the data.

Unfortunately, this means that there are no silver bullets or magic wands—pick your metaphor—to make data useful for good purposes but not for privacy invasion. This means that we always face difficult tradeoffs. Should we allow these data to be used for productive analysis, or should we prohibit or limit the analysis, because they pose too great a risk to privacy? This is a vexing problem at the heart of evidence-based policymaking. Like many others, I think we should find a balance between these competing interests. Neither should we block every proposal to study the education system, nor should we unleash educational data in every form, privacy be damned.

Faced with these difficult tradeoffs, I encourage the careful application of *friction* to systems for data analysis for evidence-based policymaking. Friction is a

[6] Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) (*hereinafter Broken Promises of Privacy*); Paul Ohm, *The Underwhelming Benefits of Big Data*, 161 U. PENN. L. REV. ONLINE 339 (2013); Paul Ohm & Scott Peppet, *What if Everything Reveals Everything?*, in BIG DATA IS NOT A MONOLITH (Cassidy Sugimoto, Michael Mattioli, and Hamid Ekbia) (2016) (*hereinafter Everything Reveals Everything*).

[7] Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125 (2015).

[8] *E.g. Broken Promises of Privacy*, *supra* note 6; *Everything Reveals Everything*, *supra* note 6.

useful metaphor because it evokes not the car careening madly down a frictionless hill, nor the car slamming into a brick wall. Like the metaphorical car, the desirable mode for privacy-protective education research sits between the extremes, *making research possible but difficult*. The right types of friction in the system can prevent wholesale privacy breaches and serve to remind researchers that these data represent the lives of vulnerable individuals, while also increasing the transparency, oversight, and accountability of the research.

Finally, privacy demands structural constraints, not merely bureaucratic controls. For example, how do we permit research into especially sensitive data without unduly risking harm to the people described in the data? The insufficient bureaucratic control approach would be to give copies of that data to researchers hoping that they will obey "rules of the road" dictating the kind of research they can do and the number and type of people with whom they can share the data. A superior structural constraints approach, in contrast, uses technology, in addition to rules of the road, to deny access to that researcher until her proposal receives thorough vetting and approval. When the research is approved, the structural constraints approach grants access to the least amount of data the researcher needs to do her work, as well as deploys other technical safeguards designed to reduce the potential for abuse or mis-sharing. An extreme version of this would be to sequester the data physically in a centralized facility cut off from the outside world, requiring the researcher to be present in the room to do the research. This extreme form of structural privacy protection has been used before, with especially sensitive individual data.[9]

3 Applying the Lessons of the Commission to Education Policy

How did I apply these four principles of privacy to my work on the Commission, and how could we apply these principles to education evidence-based policymaking going forward? To be absolutely clear, we should continue to do empirical research into the efficacy of education in this country. I am a parent of school-aged children, and I understand the need to provide parents with information about the education and welfare of their children and to provide education officials, from every teacher, to every school board member, state legislator, Department of Education official, and the members of this Congress, with meaningful information about whether our systems are working. This need is especially important because of the inequalities in our society that give rise to significant gaps in information and opportunity that disadvantage people with

[9] Centers for Disease Control and Prevention, Research Data Center: Access Modes, Centers for Disease Control and Prevention, Research Data Center: (listing datasets from the National Center for Health Statistics that are available only on site); Centers for Disease Control and Prevention, Research Data Center: On Site at an NCHS RDC, https://www.cdc.gov/rdc/b2accessmod/acs210.htm (describing security procedures for on-site access to NCHS data, including "RDC computers do not have access to the Internet nor do we allow flash drives").

fewer resources and less mobility. As we embrace empirical approaches, though, we should be mindful of the limits of social science, particularly in a field where values are as contested as they are in education.[10]

Whenever we create new research programs in education, let us learn from the privacy conclusions of the Commission's report.[11] Most importantly, the report stands for the proposition that we should not permit the centralization of massive amounts of data in the name of evidence-based policymaking. Our statutory charge instructed us to "consider whether a clearinghouse for program and survey data should be established and how to create such a clearinghouse."[12] The Commission responded to this charge forcefully by saying, in report recommendation 2-2:

> Consistent with this charge and after careful consideration of the issue, the Commission has concluded that ***a clearinghouse should not be created***. Specifically, the Commission interprets the word "clearinghouse" as connoting a data storage facility that permanently stores records from multiple databases from multiple agencies and, therefore, grows with each new data linkage. Previous panels and commissions either have come close to recommending this type of clearinghouse or did recommend one. These previous efforts provoked well-founded concerns about the potential privacy harm such a clearinghouse could raise. As further elaborated in Chapter 4, the Commission rejects the clearinghouse model in favor of [non-centralized alternative known as the National Secure Data Service].[13]

This conclusion builds on the understanding of at least some members of the Commission that when data from different agencies about millions of individuals are centralized into a single, giant database held by the federal government, trouble will likely follow. Researchers gather data with a specific research purpose in mind; those purposes fade into the background once the data are assembled into deeper pools. The result is mission creep, broken promises, defied expectations, and a lack of transparency.

Giant pools of data are also attractive honeypots for breach and misuse. Even the most sophisticated companies and government agencies in America cannot protect the sensitive information of individuals at acceptable levels.[14] We should

[10] Eloise Pasachoff, *Two Cheers for Evidence: Law, Research, and Values in Education Policymaking and Beyond,* 117 COLUM. L. REV. 1933 (2017) ("A final reason why the calls for evidence-based decisionmaking in ESSA are not going to transform education in America is that citizens are deeply divided about the underlying purpose of education.").

[11] I must emphasize that this report is a consensus document, one that represents the necessary push-and-pull of consensus-building. It is likely that other members of the Commission would not have emphasized the specific privacy-centric findings of the report that I highlight, just as I choose not to focus on the glorification of research that permeates other parts of the report.

[12] Pub. L. 114-140 § 4(b) (2016).

[13] COMMISSION REPORT, *supra* note 1, rec. 2-2, at 41 (emphasis added).

[14] Nicole Perlroth, *All 3 Billion Yahoo Accounts Were Affected by 2013 Attack*, N.Y. TIMES, Oct. 3, 2017 at B2; Tara Siegel Bernard et al., *Equifax Says Cyberattack May Have Affected 143*

worry even more about the security of hastily set up data pools, created by those whose expertise lies in the analysis rather than the protection of data.

I know that many researchers are unhappy with this line of reasoning and most likely with this conclusion of the Commission. They want access to deeper pools of data, which help give rise to more powerful inferences about the lives of the people in the data. This is, of course, precisely why such data gives rise to potentially more powerful invasions of privacy. When researchers imagine a database containing the test scores and behavioral records of millions of students, they see statistical power and profound insights, while I see millions of lives at risk of harm. Both of our viewpoints focus on the very same thing, the richness and power of the data.

This is once again an example of the intrinsic tradeoffs at the heart of this debate. Faced with this dilemma, the Commission chose to err on the side of privacy against large-scale and persistent centralization, understanding fully that this decision would necessarily place limits on the volume and type of research that will be possible.

Another key lesson of the Commission report is that we are on the cusp of significant advances in privacy enhancing technology that might permit new, more flexible forms of evidence-based policymaking with strong, cryptographically assured privacy protections. To be clear, there are no silver bullets, and none of these technologies will provide perfect privacy protection and unlimited scientific research at the same time. But they do promise the possibility of more research with fewer compromises to privacy. Two examples are "differential privacy" and "secure multi-party computation," both of which hold great promise for significant advances in the near future.[15]

The federal government is in a powerful position to spur the development of technologies like these. Congress could help the privacy research community by enacting legislation that creates pilot programs for using secure multi-party computation or differential privacy to permit a limited number of colleges to share data with decreased risk. Doing so would not only enrich education policy, it would also help advance the privacy technology state of the art. The beneficial spillover effects will go far beyond education policy.

4 THE BRIGHT PROSPECTS FOR TAKING PRIVACY SERIOUSLY IN EDUCATION POLICY

As a privacy scholar who spends most of his time thinking about government and corporate surveillance, I am struck by some of the built-in differences between either of these contexts and the education context. If harnessed wisely, these features of education policy might make it an excellent test bed for designing new

Million in the U.S., N.Y. TIMES, Sept. 7, 2017 at A1; Devlin Barrett, *U.S. Suspects Hackers in China Breached About Four Million People's Records, Officials Say*, WALL ST. J., June 5, 2015.

[15] *Id.* at 57 (sidebar on "Emerging Approaches that Enhance Privacy Protections").

systems for evidence-based policymaking that balance the need to do research with the obligation to protect privacy.

In education, the fact that so much policy devolves to state and local control is, to my mind, a feature and not a bug. The dispersed, heterogeneous, and incompatible information technology systems that hold the data about education mean that we are far from the centralized "all knowing database" that I have already argued we ought not build. This technical separation has arisen organically and historically, the product of thousands of independent choices by schools, districts, and school boards across the country, and it has been reinforced by the numerous laws that have been enacted at the state level in the past few years limiting the use and disclosure of education data for research purposes.[16]

This status quo means that education data lives apart from one another until we make a conscious and concerted choice to bring it together. With this backdrop, it is far more likely we will build new systems in small pieces that come together in fits-and-starts rather than a giant, centralized solution.[17] This is friction in action, a better result for privacy, because it increases the risk of harm in smaller steps and at a more deliberate pace.[18]

I also do not need to tell any of the members of this committee about the passionate armies of advocates who scrutinize decisions about education data intently and tirelessly. As one who laments the fact that the public rarely expresses the concern it ought to about government and corporate surveillance, I was both surprised and pleased to read the passionate and vociferous comments we received at the Commission from those who worry about student unit records and other privacy issues in education. The Commission heard and tried to respond to the concerns of these engaged participants in our process.

These three things—a diverse data system, a status quo that favors separation of data over centralization, and advocates who care about privacy—suggest that this is a context where we can thoughtfully and deliberately design new examples of evidence-based policymaking that do far more than pay lip service to privacy, one in which we can walk, not run, into new forms of data collection and analysis. I see this as a promising test bed for taking privacy seriously in social science and in policy, and I urge all who build on this test bed to take into consideration the privacy principles I have outlined such as friction and structural constraint.

[16] Future of Privacy Forum, *State Student Privacy Laws*, FERPA|SHERPA, https://ferpasherpa.org/state-laws/ (collecting state student privacy law enacted since 2013).

[17] Elana Zeide, *The Structural Consequences of Big Data-Driven Education*, 5 BIG DATA 165, 167 (2017) (describing the "highly politicized, bureaucratized, and decentralized structure of the U.S., public education system.").

[18] *See* Monica Bulger, Patrick McCormick, & Mikaela Pitcan, The Legacy of InBloom, Working Paper 02.02.2017, Data & Society (recounting the fall of InBloom, finding some positive things to say about the doomed enterprise).

5 Conclusion

In closing, we ought to pursue the ideal of evidence-based policymaking without rushing headlong into massive centralized databases that fit underneath the banner of that label but do far too little to build-in robust protections for individual privacy. The report of the Commission on Evidence-Based Policymaking and, in particular, its recommendations about privacy protection chart a way forward.