



Shine a Light: The U.S. Federal Government Should Increase Access to Educational Administrative Data

Tyler W. Watts

Teachers College, Columbia University

Drew H. Bailey

University of California, Irvine

Jade M. Jenkins

University of California, Irvine

Ana A. Whitaker

University of Delaware

Xiaolin Ou

Teachers College, Columbia University

The Institute of Educational Sciences invested substantial resources into the Statewide Longitudinal Data Systems grant program. These grants helped states convert their educational records to longitudinal datasets for research and evaluation. However, external researchers still have unreliable access to these datasets. In this commentary, we argue that the U.S. federal government should use grants to incentivize states to make educational administrative data open to researchers. The current system requires researchers to cultivate complex relationships with state and district officials, creating potential conflicts of interest that undermine the credibility of educational research. We explain how the federal government could provide a gatekeeping service that would give researchers access to statewide data systems while safeguarding student privacy. Increasing access to educational administrative data would bring transformative benefits to the field, strengthening the evidence base and helping it answer questions key to improving schools and student outcomes.

VERSION: September 2026

Suggested citation: Watts, Tyler W., Drew H. Bailey, Jade M. Jenkins, Anamarie A. Whitaker, and Xiaolin Ou. (2026). Shine a Light: The U.S. Federal Government Should Increase Access to Educational Administrative Data. (EdWorkingPaper: 26-1572). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/tcr7-hy30>

Shine a Light:
The U.S. Federal Government Should Increase Access to Educational Administrative Data

Tyler W. Watts*
Teachers College, Columbia University

Drew H. Bailey
University of California, Irvine

Jade M. Jenkins
University of California, Irvine

Ana A. Whitaker
University of Delaware

Xiaolin Ou
Teachers College, Columbia University

* *Corresponding author:*
525 120TH St.
New York, NY 10027
tw2108@tc.columbia.edu

Word count: 2491

AI Use Statement: The authors used Generative AI (Google Gemini) for copyediting, verifying legal citations, and finding specific state policy documentation. The authors also asked Gemini to assess whether the arguments presented were consistent with current federal and state standards on data access. The authors generated the article's intellectual content and wrote it. The authors assume full responsibility for the content of the manuscript.

Abstract

The Institute of Educational Sciences invested substantial resources into the Statewide Longitudinal Data Systems grant program. These grants helped states convert their educational records to longitudinal datasets for research and evaluation. However, external researchers still have unreliable access to these datasets. In this commentary, we argue that the U.S. federal government should use grants to incentivize states to make educational administrative data open to researchers. The current system requires researchers to cultivate complex relationships with state and district officials, creating potential conflicts of interest that undermine the credibility of educational research. We explain how the federal government could provide a gatekeeping service that would give researchers access to statewide data systems while safeguarding student privacy. Increasing access to educational administrative data would bring transformative benefits to the field, strengthening the evidence base and helping it answer questions key to improving schools and student outcomes.

In 2026, the Systematizing Confidence in Open Research and Evidence (SCORE) program published results from a massive effort to document reproducibility in the social sciences (Miske et al., 2026). The researchers systematically sampled articles from prestigious journals across various fields to answer a simple question: can we reproduce each article's key claims? If I have the same data and follow your analytic methods, can I reproduce your result?

SCORE's findings are concerning for the entirety of the social sciences, but they are particularly alarming for the field of Education. Only 2.9% of the articles sampled from Education journals made their data available, meaning most Education papers could not meet a basic open science standard. For the small minority of papers with available data, *none* of the claims could be precisely reproduced. Education fared worse than any other field considered in the report, including political science, economics, sociology, psychology, and business.

Many factors contribute to this problem. Indeed, a lax culture around open science dogs the field. Yet many researchers would argue that, although they would have preferred to make their data available, restrictive data-use agreements precluded them from doing so. Education researchers often work with state or district partners to obtain administrative data on student achievement, behaviors, and enrollment, and these data often come with the proviso that researchers cannot share the data. The prohibition of educational data sharing is usually justified by concerns about student confidentiality and privacy, including the Family Educational Rights and Privacy Act (FERPA)—a 20th-century federal law that provides privacy protections for educational records. Beyond strict privacy rules, researchers also navigate other complexities in state and district partnerships, including potentially misaligned incentives that constrain the findings researchers can report. This state of affairs severely hampers progress in the field,

undermining our ability to credibly answer critical questions to improve educational experiences for U.S. children.

We argue that the U.S. federal government should work to improve access to educational data from states and local education agencies (LEAs). We propose a solution that could provide researchers access to educational administrative records while guarding against political interference in the research process. We argue that our proposal would protect student privacy and generate tangible benefits for research and policymaking. We begin by reviewing the status quo of educational data access in the U.S. before discussing our proposed solutions.

The Educational Data Status Quo

Beginning in 2005, IES in the federal Department of Education (DOE) began making substantial investments in Statewide Longitudinal Data Systems (SLDS). The SLDS grant program was designed to help states turn their educational records into high-quality data systems. Schools and LEAs keep records that track basic student information, including performance (e.g., test scores and grades), enrollment, disciplinary infractions, referrals to special services (e.g., IEPs and gifted/talented programs), and student demographics. LEAs and states vary in how they maintain such records and what specific information they track, but most U.S. public education institutions collect information across these domains. As of 2023, all 50 states, Washington, D.C., and 5 U.S. territories received SLDS grants from IES to build research databases from these educational records (Meholick et al., 2023; von Zastrow & Razynska, 2025). In total, IES's investment in the SLDS program reached approximately \$930M. Beyond IES's investments, many states received additional funds to build integrated workforce and education data systems meant to connect schooling and adult employment data (von Zastrow & Razynska, 2025).

These investments have led to many improvements in public transparency and reporting. LEAs across the country now report aggregated school data to the public, allowing parents to see critical information for their public schools. At least 25 states have also set up interactive dashboards that allow members of the public to query basic statistics about educational outcomes in their state (von Zastrow & Razynska, 2025), and states regularly publish commissioned reports on various policy initiatives using these data sources. These are worthwhile applications of the SLDS databases, and IES deserves credit for these improvements.

Despite progress, external researchers still lack reliable access to these data systems. Indeed, the SLDS databases have largely supported internal research, with state and LEA employees using the data for in-house purposes, such as compliance monitoring (Meholick et al., 2023; von Zastrow & Razynska, 2025). Most states require researchers to submit some kind of proposal to a state educational agency for review. Although every state, to our knowledge, prioritizes data security and student privacy when reviewing data applications, many go well beyond that, reviewing applications for alignment with opaque state “priorities” and “objectives.” For example, Ohio grants access to its Longitudinal Data Archive based on whether the proposal is “compatible with the data-providing agency’s organizational mission and priorities.”¹ Applicants for Massachusetts’s educational data are told that “preference is given to projects that align with state education priorities.”² For Washington, D.C.’s educational data, applicants must list which district research priorities are addressed by the project.³

¹ See: <https://chrr.osu.edu/ohio-longitudinal-data-archive/access-procedures>

² See: <https://educationtocareer.data.mass.gov/stories/s/Data-Requests/8hsv-ebcv/#2.-review-research-guides-and-agendas>

³ See: https://dcps.dc.gov/sites/default/files/dc/sites/dcps/page_content/attachments/Process%20and%20Requirements%20to%20Conduct%20Research%20or%20Obtain%20Confidential%20Data%20%28May%202023%29.pdf

These types of requirements are not unusual, and from the state agency's perspective, they are entirely reasonable. Why should an agency spend state resources supporting external research projects that do not align with their policy objectives? However, this gatekeeping heavily restricts the kinds of studies that researchers might undertake. Questions that are crucial to theory or resolving areas of conflict in the literature may have no obvious tie to a given state's current educational agenda. Moreover, because access to most state education databases depends on some form of state review, researchers often avoid projects altogether that require applications for state data, given the uncertainty about whether access will ever be granted or on what timeline.

Many researchers have found that the best way to access data is by cultivating relationships with state and district authorities. These partnerships carry appeal for both researchers and practitioners. By partnering, practitioners (or policymakers) and researchers can work together to design studies that answer questions prioritized by practicing educators. Practitioners and policymakers can grant researchers highly coveted access to schools and data, and researchers can lend their expertise in designing studies that provide empirical insights for both the state/LEAs and the scientific literature. These partnerships are built on, and likely further cultivate, trust and mutual understanding between practitioners and researchers, which may, in some cases, make research more useful.

However, when researchers partner with policymakers, incentives often do not align. Researchers hope to produce publishable scientific findings that advance our understanding. Modern open science practices like pre-registration and data sharing are meant to constrain researchers' ability to shift a research project in ways that might promote conclusions that flatter practitioners, policymakers, and/or researchers but ultimately undermine the study's credibility.

In contrast, state and LEA administrators are often under significant pressure to promote initiatives that produce desirable results. A negative or null finding from a well-intentioned, collaborative research project could lead to political fallout and may have professional consequences for the administrator. In such cases, the public's interest in gaining knowledge from credible evaluations of public programs is hampered by the concern that a state or district official might have over whether a research project could harm their own careers. Relatedly, they may worry that certain research findings could lead to disinvestment from a program that they support.

When incentives are not aligned in these partnerships, conflicts of interest can arise. Researchers cannot control the outcomes of the studies that they undertake, and evaluation projects for LEA and state initiatives might uncover unexpected or disappointing findings that create friction in the partnership. Researcher careers, especially for those early in their careers or those working at non-tenure-granting institutions, depend on long-term planning for research projects. Assured access to data allows researchers to write grants, provide students with dissertation and thesis projects, and undertake the lengthy process of working toward peer-reviewed publications. For researchers involved in these projects, maintaining access to data can be career-defining, and the prospect of losing the partnership because of uncomfortable findings can create substantial stress and potentially affect which findings are shared. Over time, researchers can become incentivized to treat their data partnerships as proprietary assets, further limiting access possibilities for other researchers. These kinds of conflicts of interest would never be reported to a granting agency or journal, because no money ever exchanges hands.

In a worrying high-profile example of the risks that these partnerships carry when data is restricted for political purposes, two education policy researchers were asked to provide expert

testimony in a case about California’s response to schooling changes during the COVID-19 pandemic. However, both researchers had data use agreements with the California DOE that forbade them from releasing any data or providing testimony for other parties without permission. The California DOE tried to block one researcher’s testimony and threatened both with \$50,000 fines (Sparks, 2023). In another troubling example, researchers accused an influential scholar of refusing to provide the necessary information to replicate a claim about the efficacy of a math curriculum at an individual school. After trying to identify the school so they could attempt to verify the claim with publicly available data, the replicating researchers asserted the original claims did not hold up (Piper, 2026).

When viewed alongside the aforementioned SCORE report on reproducibility, these cases make clear that educational research must catch up to 21st-century practices for replication and reproducibility, or the field's credibility will be severely damaged. Researchers need reliable access to educational data without political prejudice—a “no strings attached” approach that puts research integrity front and center while balancing the need to maintain safeguards for student privacy and facilitating partnerships. Below, we offer a potential path forward.

The U.S. Federal Government Should Act as the Gatekeeper for Educational Data

Although IES made significant investments in SLDS grants across the country, to our knowledge, IES never explicitly required that the databases generated from these grants be made accessible to the broader educational research community. For any future SLDS grant project, IES should explicitly require that external researchers can gain access to data as a condition of accepting grant funds. Access should follow widely accepted open science standards; access can be granted based on one’s ability to meet specific security requirements, but not on subjective assessments of the research methodology, subject matter, or hypotheses.

To further facilitate this, the US DOE could act as an active gatekeeper of state student-level SLDS databases. In this model, IES would make grant funds available to states willing to place their data on a federally managed, secure server. This server could be uniquely configured for each state, and a federal agency like the National Center for Educational Statistics (NCES) could offer data curation services to states that are currently under-resourced to clean and de-identify data. NCES would maintain the data in a secure environment, and researchers would apply to NCES for access. This would relieve states of the burden of maintaining legal agreements with external researchers, a task the federal government already undertakes for secure datasets managed by the DOE and the Census Bureau. To address states' privacy concerns, the data could be kept in a secure data enclave, where researchers can interact with the data only on an encrypted federal server. Thus, researchers can run analyses while working on the server, but they can never remove the data. Instead, they may extract only aggregated statistical reports from the server—posing no legitimate risk to students' privacy.

To meet FERPA obligations, NCES would ensure all data is fully deidentified before sharing it with any researcher. States would use the current FERPA audit and evaluation exception to designate NCES as an “authorized representative” for the state, the legal exception that states already use to share data (Family Educational Rights and Privacy, 2023). Here, NCES would act on the state’s behalf to anonymize the data and curate it. Importantly, this plan follows the guidelines set forth by the Education Sciences Reform Act (2002), which specifically prohibited IES from creating a nationwide database of personally identifiable information on students. Instead of creating a centralized warehouse of identifiable data, each state’s data would be partitioned on an individual server, no linkages would need to be generated across states, and users would apply to access data for specific states. The Evidence Act of 2018, enacted by

Congress, provided protections for states that make data available to federal agencies by ensuring the data would be used only for “statistical purposes,” adding guardrails against concerns that a federal agency could misuse it (Congress, 2018).

Benefits of Making Educational Data Open

The benefits of this plan would be substantial. Although the plan would require a serious upfront investment from the DOE to build the necessary infrastructure, nearly \$1B has already been invested in state educational data systems. Our proposed plan would simply take the next step: ensuring access to these systems is broad enough to reap the benefits of the investments already made.

Increased data availability has the potential to transform the quality, rigor, and replicability of educational research. Qualified researchers would be free to pursue research questions addressing critical issues in the field without interference from state or district politics. For IES evaluation grants that examine the efficacy of particular programs, data quality would improve, and grants would become cheaper. It would become standard practice to provide NCES with the information needed to link researcher-generated data to state educational records. Researchers could then examine program efficacy using externally valid measures of student performance and achievement, and they could also promise to follow their samples longitudinally using the state data archive. Follow-up data is currently rare and expensive to generate (see Watts et al., 2019), and our plan would guard against attrition and make follow-up much cheaper. Further savings could come from reducing the need for researchers to collect their own measures of student performance and achievement. Most promising, many research projects could occur without grant funding, as students and researchers would actively seek opportunities

to work with the state databases given the low costs of secondary data analysis. In other words, if you build it, they will come.

Broader access to state educational administrative data would also strengthen RPPs. When data opacity is the norm, researchers face more pressure to sweep unwanted findings under the rug to maintain a positive relationship. However, if researchers and practitioners understand that others can verify findings by accessing the data, incentives to produce misleading reports diminish, and they can communicate more freely about uncomfortable results. With open data, the RPP's research integrity will improve, strengthening outcomes for schools and children in the long term.

Conclusion

Although IES oversaw crucial innovations that improved the credibility of educational science, Education still trails other fields in data access and availability (Miske et al., 2026). In the next phase of IES, the Institute should honor its previous investments in building state data systems by helping researchers access them without prejudice. Doing so would significantly improve the field's credibility and help researchers provide reliable answers to questions that could strengthen outcomes for students across the country.

References

- Congress, U. S. (2018). Foundations for evidence-based policymaking act of 2018. *Public Law, 115*, 435.
- Education Sciences Reform Act of 2002, 20 U.S.C. § 9572(a) (2002).
<https://uscode.house.gov/view.xhtml?edition=prelim&num=0&req=granuleid%3AUSC-prelim-title20-section9572>
- Family Educational Rights and Privacy, 34 C.F.R. § 99.31 (2023).
<https://www.ecfr.gov/current/title-34/subtitle-A/part-99/subpart-D/section-99.31>
- Meholick, S., Honey, R., & LaTurner, J. (2023). Profile of State Data Capacity in 2019 and 2020: Statewide Longitudinal Data Systems (SLDS) Survey Descriptive Statistics. Stats in Brief. NCES 2022-051. *National Center for Education Statistics*.
- Miske, O., Abatayo, A. L., Daley, M., Dirzo, M., Fox, N., Haber, N., Hahn, K. M., Struhl, M. K., Mawhinney, B., Silverstein, P., Stankov, T., Tyner, A. H., Adamkovič, M., Alzahawi, S., Anafinova, S., Awtrey, E., Axxe, E., Bailey, J., Bakker, B. N., ... Errington, T. M. (2026). Investigating the reproducibility of the social and behavioural sciences. *Nature*, 652(8108), 126–134. <https://doi.org/10.1038/s41586-026-10203-5>
- Piper, K. (2026, April 17). *Education research is weak and sloppy. Why?*
<https://www.theargumentmag.com/p/education-research-is-weak-and-sloppy>
- Sparks, S. D. (2023, August 23). Lawsuit Spotlights California's Restrictions on Researchers' Use of Its Education Data. *Education Week*. <https://www.edweek.org/policy-politics/lawsuit-spotlights-californias-restrictions-on-researchers-use-of-its-education-data/2023/08>

von Zastrow, C., & Razynska, K. (2025, November 17). *Sustaining state longitudinal data systems*. Education Commission of the States. Retrieved from:

<https://www.ecs.org/sustaining-state-longitudinal-data-systems-slds/>

Watts, T. W., Bailey, D. H., & Li, C. (2019). Aiming Further: Addressing the Need for High-Quality Longitudinal Research in Education. *Journal of Research on Educational Effectiveness*, 12(4), 648–658. <https://doi.org/10.1080/19345747.2019.1644692>