Bridging the Gaps in Education Data

By Jon Fullerton

October 2021

Key Points

- Over the past two decades, education underwent a "big data" revolution as states began tracking individual student performance and interim assessments and educational software allowed for a greater granularity of data on students, teachers, and schools.
- Despite this plethora of new data, considerable gaps in data on early childhood education, school spending, student programs and interventions, and postsecondary outcomes remain.
- Dissatisfaction with education data will never fully disappear due to technical gaps between what policymakers and researchers would like to measure and what can be measured, as well as normative disagreements about the ends of education.
- Policymakers should focus on closing the gaps they can while also recognizing the technical and normative constraints on measuring outcomes.

During the past 20 years, as education moved into the world of big data, big dreams abounded. In this new world, it was hoped that educators would have the data at their fingertips to easily identify student strengths and areas in which they needed extra help. Administrators would use predictive analytics to identify which students were most likely to struggle and target supports appropriately. Researchers would produce catalogues of interventions that would allow districts to know what works and what doesn't and therefore not waste time and resources on the latter.

Two revolutions in the early years of this century made such changes conceivable. First, with the advent of the No Child Left Behind (NCLB) Act in 2002, states were required to track individual students' academic performance in grades three through eight and other outcomes at regular intervals. Before this, students were tested only every few grades, making it impossible to track an individual student's growth.

Around the same time, education agencies began to adopt modern student information systems and enterprise resource planning systems that made it possible to collect and connect data across domains in K–12 education. One could now connect student attendance and grades to state achievement tests, teacher records from HR, and payroll data. Third-party advocacy groups such as the Data Quality Campaign and the Ed-Fi Alliance urged agencies to adopt standard data identifiers and structures that would allow for both interoperability across disparate data systems and the tracking of students over time. The Department of Education's Statewide Longitudinal Data Systems grants helped fund the creation of these new data systems

and support their usage. Now one could look at individual student growth from year to year and at school or teacher impacts on that growth.

Over time, numerous states also began to connect current and former student data to data from other systems outside the K–12 system, such as higher education enrollment and wage data from state unemployment systems. This has allowed new insights into who is attending, and succeeding in, college; what work and wage outcomes are for students coming out of different educational pathways; and much more. Researchers and policymakers can now understand better the long-term trajectories of public school students without special data collection efforts. Partly because of this, schools are no longer seen simply as responsible for graduating kids but for ensuring that their students are prepared for college and career.

A second revolution came in terms of the granularity of data that is collected on students, teachers, and schools. Partly as a reaction to the accountability pressure brought by NCLB, a new interim assessment industry sprang up. At least in theory, teachers could identify specific concepts students were struggling with and gaps in their learning using detailed data from these online assessments. The assessments' data reporting tools could also provide department heads, principals, and district administrators new insights into how classes and schools as a whole were doing and where the system was struggling—within an instructionally relevant time frame.

To help students close the gaps that the state and interim assessments highlighted, many schools began to use new educational software to supplement teacher-led instruction. Such software can track not only student progress but also student (and teacher) effort—how long students spend on the program, how many tries it takes for a student to "master" a topic, and where students tend to get stuck. Students and teachers create rich data streams as a by-product of working with interim assessments and software. However, unlike the longitudinal data mentioned above, much of these data live in the systems of individual providers and are not standardized across districts. It can thus be much more difficult for managers at higher levels of the system to access or for analysts to mine and explore for insights.

These two revolutions have changed school systems in myriad, generally positive ways. My part of the ecosystem—education research—was transformed by the availability of regular outcome and administrative data on students and teachers. Treatment and comparison groups could now be constructed and studied without multimillion-dollar data collection efforts. Schools, districts, and colleges have used the new data to set up early warning systems to identify which students are likely or beginning to struggle before they fall offtrack and to monitor or nudge students through some crucial hurdles on the way to college.

On the system policy and management side, education agencies can now contemplate measuring the outcomes of interventions in trials before rolling out an intervention to the whole system. Data can be used to not just determine what went right or wrong but guide future decisions.

Despite this plethora of new data and the new uses to which they have been put, dissatisfaction with the data available remains, and data dreams continue unfulfilled. Why?

In this report, I argue that there are two distinct answers to this question. First, there are several areas in which we ought to be collecting more data and in which the information revolution within education has not kept pace. These are solvable, if not necessarily simple, problems. Second, even if we bridge these data gaps, dissatisfaction will continue to stem from deeper roots. There are and will continue to be large technical gaps between what we would like to measure and what we can collect at a reasonable cost. Indeed, some of these technical gaps cannot be closed at any cost.

In addition, there are important normative disagreements about what educational outcomes we should be measuring. When there is not agreement on the goals of education, it is hard to collect data that measure progress toward them. Ignoring these normative disagreements can undermine the legitimacy of the data and the measures we do have.

While the measurement and data situation in education has improved dramatically (and, I hope, will continue to improve), these last two issues will never go away. As a result, we have to keep expectations of what "better" data can realistically offer as we tinker toward more robust and useful metrics.

Gaps Remain

Despite the leaps in data availability that have taken place over the past 15 years, some important and surprising gaps still undermine schools' ability to manage and policymakers' ability to identify challenges and look at the effects of their policies. Because of these data gaps, insight into educational institutions remains hazier than necessary.

Early Childhood Education Data. Since the first findings on the outcomes of the Perry Preschool Project were released in the early 1970s, early childhood educational experiences have been considered crucial to later educational success. Indeed, there has been great emphasis over the past 20 years—and billions of dollars of funding—on providing richer- and higher-quality preschool experiences to more children, especially socioeconomically disadvantaged children. Numerous states have either increased or made universal the availability of preschool to parents who could not otherwise afford it.

However, the data that would allow the public, state policymakers, and schools the ability to understand the impact of these investments have not kept pace.

It is not that no data exist. Most states have some form of licensing for day care and preschool providers, and 43 states have implemented quality measures as part of the licensing process.[1] So, it is possible to know who is offering a licensed preschool and (often) whether those preschools meet certain "quality" benchmarks. However, in many states, the data end there. Which children actually attend which programs is unknown, much less information about what skills children emerging from the programs have and whether they are at risk for possible learning challenges in the future.

A key reason for this data gap is that, unlike K–12, preschool and day care offerings are highly decentralized and often live outside the traditional district structure. Preschools range from large, multistate private organizations to preschool classes offered by local districts to preschools with little "academic" content offered by small, local providers to their neighbors. Because of this decentralization, accessible records of where, or whether, a child attended preschool are not often available.

School districts know, of course, whether a child went to the district's own preschool program, and often districts might know who attended Head Start programs or state offerings. But this information covers only a fraction of kids. In a recent study in Massachusetts, only 20 percent of sampled 4-year-olds participated in public school preschool or Head Start, with about 50 percent participating in informal care.[2]

How did the authors of the study gather information about the informal sector? Well, it wasn't from the state. Instead, the authors had to set up their own tracking systems for a sample of children and then individually follow the children for the study. This is not a replicable approach to support continuous insight into the pre-K system.

In addition, even if elementary schools knew where kids were before age 5, little data exist on the quality of teaching or on what skills children learn before kindergarten. A recent survey of early-assessment use across states notes both the paucity of practical, high-quality assessments for this age group and the fragmentation and lack of coordination across providers of which assessments to use.[3]

Because of this extreme decentralization, district leaders and policymakers lack the information needed to understand what support children coming from different early education backgrounds could use, and parents have little data on which to base their decisions of where to send their kids for preschool. In addition, there is little ability to provide feedback to early education providers for increasing their students' success in elementary school. Given the funds pouring into expanding preschool offerings, fixing this data challenge is crucial. However, given the decentralization of the sector, states will need to take the lead to get comparable, clean data and create the new data systems all providers can use.

School Spending Data. The fairness of state school finance systems has been debated for more than 50 years, with every state making major changes over this time in how money is raised and distributed to school systems. At the same time, debates about the efficiency (or inefficiency) of school spending have raged. Yet, amazingly, only with the passage of the Every Student Succeeds Act

(ESSA) in 2015 have states been required to report publicly per-student spending at the school level (as opposed to the district level)—allowing the public and legislators to see relative spending across not just districts but also schools in districts. This reporting requirement has just taken effect in 2020.

What are the implications of this? For the first time, it will become apparent whether resources provided by states are reaching their intended target students. Also, for the first time, it will be easy to see variations in salary spending across schools. In the past, districts tended to use "average salary" to calculate school-level spending. This meant that when reporting spending, districts simply multiplied the average teacher salary across the district by the number of teachers in any given school to determine instructional spending—regardless of whether a school was struggling with turnover and, as a result, had many relatively cheap junior teachers or was filled with much more experienced (and thus much more expensive) veteran teachers.

This allowed for large, hidden differences in spending per student to accrue across schools.[4] Happily, under ESSA such differences can no longer be swept under the rug. As spending differences across schools are made public for the first time, we should expect some long-needed and difficult conversations to emerge about district priorities and relative school efficiency. However, while this is a tremendous step forward, there is still a long way to go before we have reasonable transparency in school-level spending and the ability to connect spending patterns to results.

First, outside teacher salaries, how spending gets attributed to schools is not necessarily consistent across districts within a state. For instance, issues can arise when some districts budget a service (e.g., counseling) centrally while other districts budget the same service to schools.

How the first district decides to allocate the cost of the program will matter. For instance, a district could choose to allocate costs on a per-student basis—assuming that, on average, students will each need the same amount of service. Another district could allocate counselors out on a per-school basis—assigning, say, half a full-time counselor position to each school. A third district might allocate costs on a usage basis, with schools being charged for the actual amounts of time counselors spend with their students.

Such district choices will affect both the relative differences in spending across schools within districts and relative spending in schools across districts. In addition, different states have their own rules for how districts should apportion centrally budgeted spending. This will further hinder the ability to compare schools across states. (To tackle at least a portion of this challenge, several states have come together to develop Interstate Financial Reporting standards, in which states agree on a standardized reporting of some data points.)[5]

There is still a long way to go before we have reasonable transparency in school-level spending.

But there is a second, more important gap in the available spending data. While new per-pupil school-spending data allow variation in spending across schools to be seen, the reporting requirements are in very broad categories—insufficient to understand the return on investment of different programs or spending choices across schools. In other words, one cannot discern how much spending goes toward extra academic help for English learners versus tutoring versus professional development for teachers.

Given the constraints of federalism, it may be impractical to design a single financial coding structure in enough detail to capture this type of information nationally. However, many states[6] do not collect and publish program data in a standardized way with sufficient detail to allow for understanding what is actually being spent on programs—essential to understanding their cost-effectiveness. Without these data, it is hard to see how schools can learn from one another's resource decisions or even whether certain spending patterns result in better outcomes.

Student Program and Intervention Data. This points to a second large gap in the K–12 data: Not only do we not know how much is spent on given

interventions at schools, states and many districts don't know which students are receiving which interventions. If no one beyond the school knows who received high-dosage, intensive tutoring, it can be guaranteed that systems will not be able to determine whether tutoring worked, the cost of tutoring relative to student growth (as opposed to simply number of students served), and whether tutoring is more or less cost-effective than other interventions for catching kids up.

Individual school leaders, of course, do (or should) know which students are receiving interventions. The problem is a data gap among schools, districts, and states. Like most data gaps, this exists for a reason.

Many of the interventions students receive are school based, and almost all are district based. In other words, schools and districts will vary in whether they use ST Math, DreamBox Learning, or another program for supplemental online math instruction. Likewise, districts that provide tutoring to students may use tutors from a national organization such as Saga Education, or they may rely on much smaller, local mom-and-pop shops to provide this service. As a result, if one captured simply the names of the service providers, one could end up with a long list of providers with only a few locales using any given provider.

In addition, analyzing the effectiveness of a given tutoring intervention would require substantial information on what services are actually being provided. In other words, is a student being tutored in math or history? Is the program during the school day or after school? How many minutes a week does the student receive tutoring? Does the student even attend the sessions? Without answering these questions, the long list of providers will not provide much insight into what students received. Districts, if they have these data at all, do not have them standardized across schools.

So, for system-wide learning, two problems need to be solved. First, can the long list of individual providers be shortened into a tractable number of similar interventions? Second, can data collected about interventions be standardized such that characteristics of interventions aimed at similar problems can be identified? In short, there should be a typology that allows schools to select and track program and intervention participation for students. For ease of tracking and reporting, this typology should be integrated into student information systems as much as possible.

So, a student receiving high-dosage tutoring might get tagged with "high-dosage tutoring, [Provider Name], math, two hours per week, in-school, in-person tutors." This would allow policymakers to get a bird's-eye view of what strategies are being used and allow researchers to begin to identify which strategies are effective for which kids. The key is standardizing what data get collected across schools and districts.

Such standardization is unlikely to arise at scale from bottom-up processes but will rely on state or national standards and reporting requirements to get going. With the passage of the American Rescue Plan and the two earlier COVID-19 recovery bills, almost $200 billion are being pumped into school systems, much of which are intended to help kids catch up from missed learning opportunities over the past year. This would be an excellent time to start collecting these data.[7] Without requiring some reporting back on which students are receiving which services, we will have difficulty learning what worked—or even knowing where this unprecedented amount of money went—at the end of the investment.[8]

Postsecondary Outcomes Data. Just as there is a significant data gap at the intake year for the K–12 system, so there is a gap after students exit the system.

The situation is much better than it was 15 years ago. For many years, schools and districts with interest have been able to monitor whether and where their graduates have attended college. School systems have been able to connect college attendance data, provided either by their state system for institutions of higher education (IHEs) or by a private company called the National Student Clearinghouse, to their own data on K–12 graduates. This has allowed schools to better monitor their program success (whether students make it to college, persist, and graduate); uncover issues schools were previously blind to such as "summer melt" (in which high school seniors who have been accepted to college and plan to attend in the spring don't show up in the fall); and evaluate numerous policy initiatives aimed at getting more students into college.

However, most school systems can know little beyond whether their graduates go to college and, eventually, graduate. This is despite an increasing emphasis on students achieving success through pathways other than college.

The problem is the data available on these pathways are weak to nonexistent. While some states can track whether former students are employed in the state, this still leaves great gaps in knowledge about student success. Students who join the military or leave the state for another job (both of which one might count as successful postsecondary outcomes) simply do not show up in state unemployment insurance data and thus are counted the same as students who do not get a job.

The data for students that states can track generally do not go beyond wages; thus, it is hard to tell if a graduate is in a low-paying, low-skill job or an internship or other skill-building program that, while initially low paying, is designed to lead to higher-skilled jobs in the future.

The data for students that states can track generally do not go beyond wages.

In good news, through the Coleridge Institute at New York University, some states have begun establishing secure data sharing arrangements that allow analysis across states to occur. For the first time, states that share a substantial amount of their labor pool (e.g., Kentucky and Ohio) can track and explore their students' success after high school.[9] Unfortunately, most states still do not have a good system for recording internships and other such programs and are thus not able to determine which of these make a difference.

Why Dissatisfaction with Data Will Continue

Filling the data gaps discussed above would allow for more informed policymaking and provide systems and schools with actionable information to guide their educational strategies. However, even if states and schools implement the systems needed to collect, distribute, and use all the above data, we will still find ourselves dissatisfied with the information we have. Crucial pieces of data will continue to be missing, the data that are collected will be dismissed as insufficient or misconceived, and K–12 education will continue to be seen as data poor.

While the four areas discussed above for improving data collection are not exhaustive, no matter what data we collect, we have to understand that they will always come up short. Why is this?

Technical Challenges. First, there are technical challenges. By this I mean that measures, particularly of educational outcomes, almost never capture the richness of what we want to measure efficiently and in time to be useful. Using some of the most common measures of educational outcomes—grades, standardized test scores, and degrees—we can see the issues.

Grades are not particularly reliable in demonstrating students' actual skills. Some teachers, and indeed schools, expect more or less for a particular grade, resulting in transcripts that are not comparable and that may not reveal much about a student's true level of preparation—at least as compared to students in other classrooms and schools.

On the other hand, the most common standardized tests (state assessments, the SAT, the ACT, etc.), which are designed for reliability and comparability, are accused of (1) not assessing true and "deep" understanding of the subjects they cover and (2) omitting the important skills and competencies, such as creative thinking and problem-solving. As a result, it is feared that these tests merely measure test prep and the critical skills needed for the workplace—such as the ability to solve novel and complex problems and work with others—are not measured at all.[10]

Finally, using degrees as markers of accomplishment has several flaws. Similar to grades, what a degree (high school, college, or other) means regarding skills and knowledge is not clear. At the secondary level, some states have attempted to moderate this issue with high school exit exams required for graduation, reintroducing the issues discussed above with standardized tests.

However, a more important flaw with degrees as markers of accomplishment has recently been getting more attention. Degree requirements for

employment exclude those who have or could easily develop the required skills but did not go to college.[11] In addition, the direct and opportunity costs of a college degree are higher than many can pay—and can involve irrelevant coursework that does not directly help the students achieve their employment and life goals.

And these are the easiest outcomes to obtain. Schools can measure reading, math, and writing skills fairly well. They can be tested directly and close to the time during which the skills are supposed to be acquired. Despite the possibilities of cheating and poorly designed tests, it is still generally hard for students to solve a math problem if they really don't understand the math.

This is different from trying to measure other potential educational outcomes of import (e.g., work habits or democratic competence). Surveys used to measure such things are rife with problems—particularly if such surveys are being used for accountability for schools or students. Respondents' answers can be influenced both by their reference groups (e.g., who they are comparing themselves to when they answer that they do, or do not, have "grit"[12]) and what they think those doing the surveying want to see. Even if these issues are solved, there is a gap between what people describe as their character and how they actually act in real life.

Now, many of the above measures could in theory be replaced with more "authentic" measures. Portfolios scored in by teams of expert teachers could replace grades and exams for demonstrating student competence. Alternatively, students could be given long-term tasks that specifically test their grit. However, the staff and student time and resources needed to execute such data gathering would be substantial—and quite possibly well beyond what is affordable or practical for state and local governments.

Even assuming that resources were not constrained, there remains an unsolvable technical challenge—time. Many of the outcomes we are most interested in are temporally distant from any immediate educational actions. Using future wages to understand a school's impact on its student success has a lag of 10 to 20 years—and there is no way to shorten this time.

While this information can help develop policies in the long term, it will not help parents' or principals' short- or medium-term decision-making. Outcome measures that will be known in five, 10, or 20 years are useless for incenting or empowering those who are running schools now. This temporal issue cannot be solved—and, as a result, education leaders and the public cannot know with certainty how today's education will affect students as adults.

Normative Disagreement About the Goals of Education. These technical challenges, however, assume we have a clear idea of what we are trying to measure. Current discussions of metrics in education tend to focus on their availability, granularity, and the technical properties (e.g., reliability and validity). But often missed in these debates is that we don't agree, at least in any deep way, on the specifics about what education is for and what we are even trying to measure. Without a consensus on the ends of education, any measurement system designed to support achieving those ends will fail.

Schools are asked to educate students in many ways: develop basic academic skills in reading, writing, math, science, and social studies; teach students problem-solving; develop effective work habits and ability to work with others; help students maintain and develop positive mental habits; provide the background knowledge and support needed to be a democratic citizen; shape students' moral character; and provide students with skills and habits needed to live successful lives.

Often missed in these debates is that we don't agree, at least in any deep way, on the specifics about what education is for and what we are even trying to measure.

While such a list may not seem particularly contentious, what each educational charge entails is. Once one delves beyond the most basic platitudes, sharp disagreements emerge.

How do we define "democratic competence"? Does a good citizen know the history of the country (which interpretation of history?) and the basics of

how the government works? Does a citizen actively engage with their community, building social capital through engagement in civil society and participating in local self-government? Does a good citizen protest, or even revolt, against unjust laws and authorities? And what concept of justice should guide these answers?

Are there higher ends to which education should seek to align students? Should schools be designed to maximize the development of adult autonomy—opening up children to as many different possibilities of arguably good lives as possible—or should schools impel children toward certain preferred versions of the good life? There is no agreement here.

Perhaps we can ignore these questions and just look at students' life outcomes. However, normative issues loom large here as well.

For instance, if we measure employment outcomes, what kinds of jobs should be valued? It is easiest to measure former students' wages. But if wages are the outcome measure of success in education, we have adopted a not too subtle value judgment that investment bankers are simply more successful than (and thus preferred to) elementary school teachers. Many will reasonably object. On the other hand, happiness experienced or social good done by members of a profession would be challenging to measure, even if there was broad agreement on what "social good" is.

Because education measurement and data discussions often avoid normative conflicts around ends, the measures that result tend to be thin and unsatisfying. Measures gravitate toward the most easily available or the lowest common denominator, which are less controversial but also less reflective of the many and conflicting higher goals of education. Wages and voting become the outcome measures of a successful education, and all (except for the odd economist or political scientist) are left with a feeling that the measures are missing the target.

Conclusion

While we can and should certainly extend what we measure, we need to recognize that the data will always be limited and cannot dig us out of normative disagreements. As a result, our big data dreams may never be fully realized.

That said, there has been tremendous progress over the past 20 years in data collection and use. As discussed above, we can now see challenges that were previously invisible (e.g., summer melt, disparities in school-level spending, and student assignment to teachers). We can also monitor outcomes of programs and have the comparison data to determine impacts on the imperfect measures we do have.

We have an obligation to continue this work and improve the data we collect. But we must also maintain an appropriate sense of humility and realize that data will not answer all our questions or make our core disagreements go away. We may not be able to get all the data we want, but we can at least be closer to having the numbers we need.

About the Author

Jon Fullerton is the executive director of the Center for Education Policy Research at Harvard University.

Notes

1. Lynn Olson and Brooke LePage, "Tough Test: The Nation's Troubled Early Learning Assessment Landscape," FutureEd, March 1, 2021, https://www.future-ed.org/tough-test-the-nations-troubled-early-learning-assessment-landscape/.
2. Stephanie M. Jones et al., "Exploring the Role of Quality in a Population Study of Early Education and Care," *Early Childhood Research Quarterly* 53 (2020): 551–70, https://www.sciencedirect.com/science/article/abs/pii/S0885200620300764.
3. Olson and LePage, "Tough Test."
4. For one attempt to get at hidden salary gaps that result from this, see Raegen T. Miller, *Comparable, Schmomparable: Evidence of Inequity in the Allocation of Funds for Teacher Salary Within California's Public School Districts*, Center for American Progress, May 2010, https://cdn.americanprogress.org/wp-content/uploads/issues/2010/05/pdf/comparable_schmomparable.pdf.
5. Edunomics Lab, *Interstate Financial Reporting: Making the Most of School-Level Per-Student Spending Data*, March 2018, https://edunomicslab.org/wp-content/uploads/2018/03/Interstate-Financial-Reporting_FINAL-V2.pdf.

6. Rhode Island does this, publishing detailed, line-item expenditures for all its districts. See Rhode Island Department of Education, "RIDE's Uniform Chart of Accounts," https://www.ride.ri.gov/FundingFinance/SchoolDistrictFinancialData/UniformChartofAccounts.aspx. Even here, though, there are challenges as most accounting systems do not make it easy to see all expenses of a program or how many are served. On the limitations of standard accounting systems, see Allan Odden et al., "Defining School-Level Expenditure Structures That Reflect Educational Strategies," *Journal of Education Finance* 28, no. 3 (Winter 2003): 323–56, https://www.jstor.org/stable/40704173.

7. Thomas J. Kane, "How We Can Identify What Students Need to Catch Up on Learning," *Education Week*, April 27, 2021, https://www.edweek.org/policy-politics/opinion-how-we-can-identify-what-students-need-to-catch-up-on-learning/2021/04.

8. There are some ongoing efforts to collect this type of data. For instance, the Center for Education Policy Research at Harvard University (where I am executive director), in partnership with the American Institutes for Research and NWEA, is working with a set of districts to collect this type of data and measure what seems to be working across participants. However, the number of districts that can be included in this work is relatively small compared to what would be possible if states (or the federal government) took the lead on collecting intervention data.

9. See, for instance, Kentucky Center for Statistics, "Work Ready Communities," March 9, 2021, https://kystats.ky.gov/Reports/Tableau/2021_WRC.

10. There is also a worry that assessments are racially biased, thus distorting the comparability of scores. This, however, is beyond the scope of this report.

11. Peter Q. Blair et al., "Searching for Stars: Work Experience as a Job Market Signal for Workers Without Bachelor's Degrees" (working paper, National Bureau of Economic Research, Cambridge, MA, March 2020), https://www.nber.org/papers/w26844.

12. Martin R. West et al., "Promise and Paradox: Measuring Students' Non-Cognitive Skills and the Impact of Schooling," *Educational Evaluation and Policy Analysis* 38, no. 1 (March 2016): 148–70, https://journals.sagepub.com/doi/full/10.3102/0162373715597298.