Link It, Open It, Use It

Changing How Education Data Are Used to Generate Ideas

Jack Rossiter

Data-driven decision-making is in the spotlight in 2020, with the public expecting data to guide government choices during ongoing emergencies, including COVID-19. Within education systems, leaders want to know what the data can tell them about when to reopen schools, how to prevent learning loss, how to gauge dropout risk, how to encourage re-enrolment, where to deploy teachers to manage class sizes, and how to meet the needs of students when they return to schools.

In response, countries have scrambled to setup systems that track health and economic indicators and monitor equity and mobility issues that may have been caused by government action. Yet as ministries of education plan their recovery, many rely on data systems which fail to provide the information needed to target attention; and as planners seek to learn from past emergencies in similar contexts, they are finding that public data on basic indicators like re-enrolment and teacher supply do not exist. Beyond urgent needs, our current crisis is drawing attention to long-standing flaws in education data systems.

In this note, I discuss a new approach to how national administrative education data—records of school census, public exams, school inspection, teacher payroll, and other operational matters, collected on routine basis—are integrated, shared, and used to generate knowledge. Drawing on examples from low- and middle-income countries, I demonstrate (a) how integrating and making available administrative data will deliver relevant policy and operational insights; (b) how this approach can engage individuals with the skills and incentives to solve data and policy problems with ministries of education; and (c) how increasing data use can be the fastest path to improving what is collected. I present Open Data for Education System Analysis as a strategy to change the way education data are used to facilitate better analysis and evidence for education.

THE PROBLEM

Researchers often produce high-quality evidence, but it is expensive to generate, tends to be small-scale and insulated from political economy challenges, and can miss the questions that are important to education policymakers. At the same time, governments are making large and regular investments in gathering data, often driven by top-down requirements from development partners. But they get very little in return: despite substantial government and partner support to statistical capacity, the

skills and systems needed for integration are unavailable in most education ministries. Data remain disjointed, and severe limits on who can access them internally and externally mean that they are barely used. This matters: a growing body of evidence[1] suggests that growth and development are slower where access to data and research materials are limited; and while some findings are transferable, analysis using local data is often a better guide to policy than experimental estimates from a different context.[2] This suggests a benefit to countries from maintaining their education administrative data and letting people use it.[3]

THE SOLUTION

Governments should *integrate* and open access to national administrative data, then support analysts to use it. Annual school census data are useful, but they have far fewer uses than merged school census, public exams, and teacher payroll data. Once integrated, these data can answer questions that are impossible to answer with survey data alone, and they are particularly well-suited to policy evaluation at a scale that matters to decision makers. Integrated administrative data are comprehensive and representative, covering whole populations in which it is possible to detect modest but meaningful relationships with precision, even in sub-groups. Their use allows officials to identify, in real-time, where existing policies require adjustment. This is not about creating a top-down, command and control education system. Instead, by opening access as a strategy, governments can leverage their data to attract individuals and organizations with the skills to support data integration and generate relevant ideas.

THE BOTTOM LINE

If we want to see evidence used to drive educational improvements, then governments and government data should be our priority. A central message of the 2021 World Development Report will be that "much of the value of data is untapped, waiting to be realized." Experience in 2020 has revealed how urgently data systems and practices need to transform so that governments can use the data that they collect and hold. Open Data for Education System Analysis is a strategy to facilitate analysis of high social value and—ultimately—of better evidence for educational development.

THE PROBLEM: HUGE DEMAND FOR INSIGHTS AND LOTS OF DATA …
THAT ARE BARELY USED[4]

Governments generate the vast majority of data on education. Line ministries routinely produce huge datasets that encompass nearly entire populations.[5] As a result, there is an abundance of education administrative data related to services provided, the people who access those services, and the outcomes of those services.
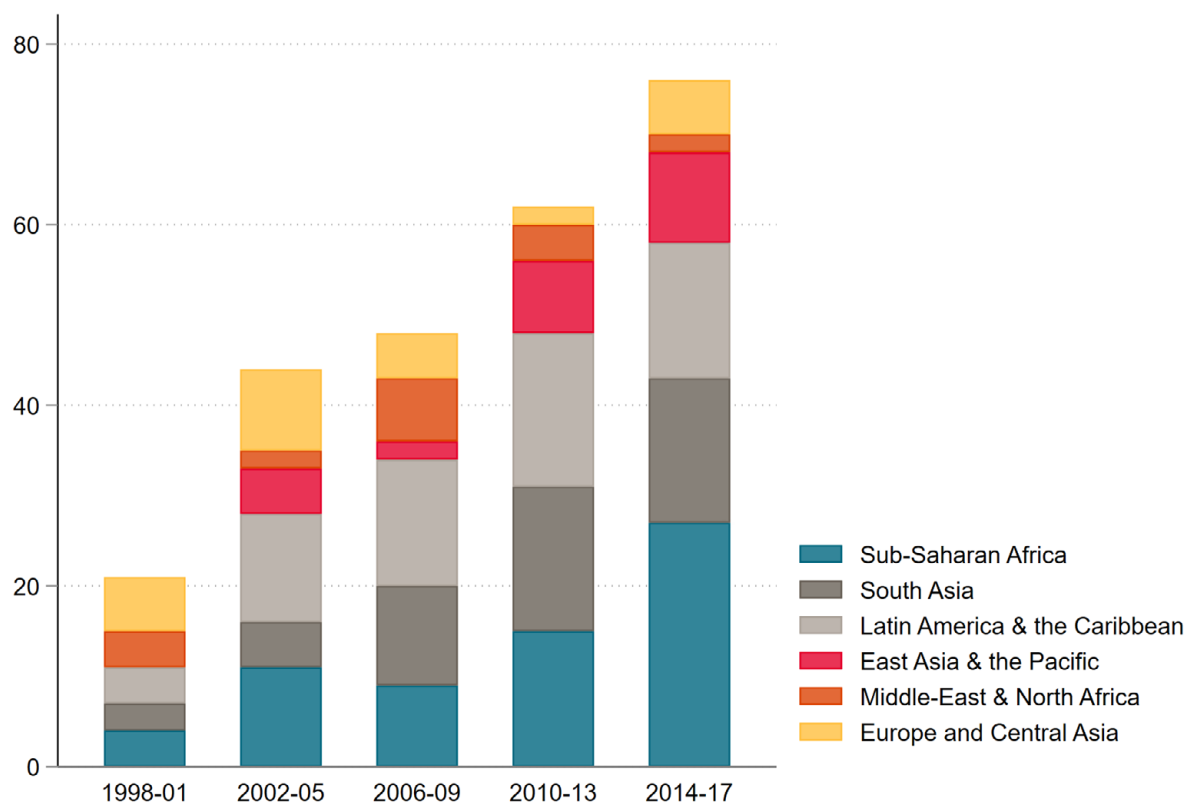
Generating population-level data requires substantial resources. An estimate for 77 of the world's lower-income countries[6] put the annual cost of utilizing survey, census, and administrative data to monitor the full set of Sustainable Development Goals at around $1 billion. Around 10 percent of this was for improving and implementing Education Management Information Systems' (EMIS) annual data collection and analysis—roughly US$1.0 to $1.5 million per country per year.[7]

National administrative data are comprehensive and representative—and this matters to policymakers. These data, by virtue of their population-level nature and the frequency of observation, allow

analysts to follow districts, schools, and sometimes students over time. They provide the level of coverage and disaggregation that officials demand for policy applications. In 2017, AidData fielded two surveys of policymakers in low- and middle-income countries: *Listening to Leaders* and the *Education Snap Poll*.[8] Participants were asked what information they value the most when making decisions and why. Overwhelmingly, education decision-makers prioritize their administrative data systems as sources of timely, disaggregated, and locally relevant information, to inform resourcing and planning processes.

International support to statistical capacity is vast and continues to grow. Donor commitments to statistics and statistical capacity doubled over the decade to 2017, with the World Bank leading in aid to statistics.[9] In the education sector, between 1998 and 2017 the World Bank funded 279 projects with EMIS components in 98 countries. The number of new projects continues to rise in each period (Figure 1) and the country count understates the full scale of the efforts, as there are EMIS regional projects as well.[10] A 2017 study showed that the average cost of the World Bank's EMIS development and strengthening activities ranged between US$1 million and $7 million per project.[11]

Figure 1. Number of new World Bank operations with EMIS activities, by region and start date
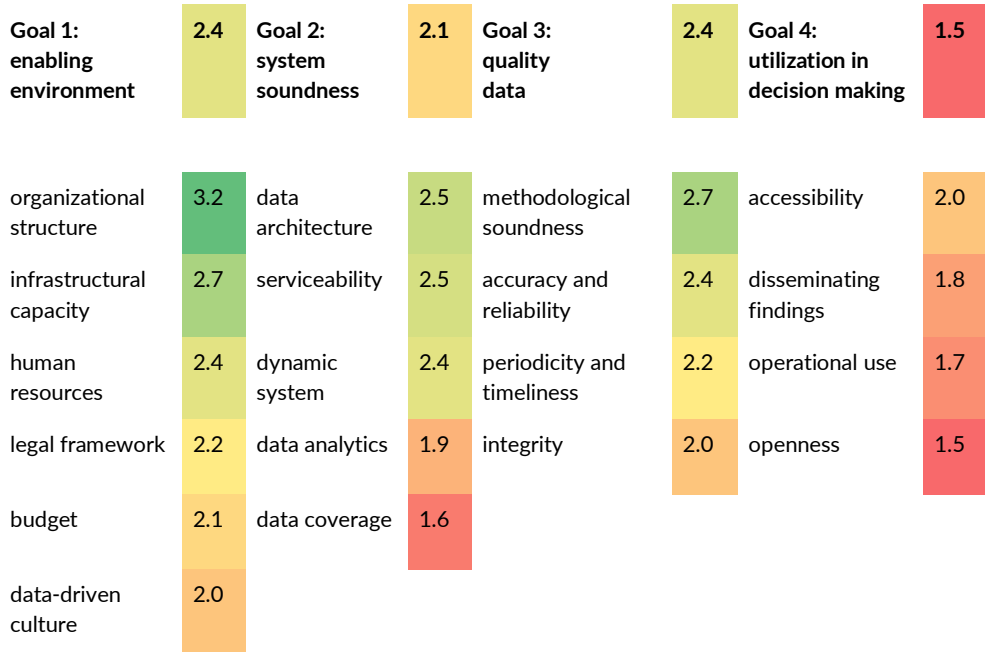

Source: EdStats Database: http://datatopics.worldbank.org/education/wDashboard/dqemis

CHANGING HOW EDUCATION DATA ARE USED TO GENERATE IDEAS

Despite all the activity, within ministries of education, capacity to produce data vastly outstrips ability to use it. From research and reviews into EMIS activities since 1998, there are several examples of projects improving data collection but failing to improve data use:

- Timor-Leste has introduced EMIS infrastructure and applications and improved data collection and processing, but "no capacity is found as yet within the EMIS for exploitation of the database for analytical work … [or] to underpin policy making." [12]

- Ghana's Ministry of Education geared its systems towards producing, compiling, and generating a statistics yearbook but after release "lack of utilization of data was a major problem."[13]

- Researchers suggest that India's EMIS reforms have changed the role of local officials from identifying, deliberating, and communicating challenges faced by schools into "merely collecting data points and feeding checklists and formats to the higher authorities in as little time as possible."[14]

- In Ethiopia, education and social protection programs have included regular and large investments in the EMIS system since 2009.[15] Such is the lack of impact on statistical capacity that a decade later, the latest program finds "deficient or inadequate systems for data collection and analysis" and includes payments of $2 million and $3 million if the education ministry publishes an annual statistical abstract and a single analytical report, respectively.[16]

Periodic assessments of country EMIS are also carried out under the SABER initiative, a World Bank initiative to produce comparative data on education policies and institutions. SABER assessments show similar limits to data access and use (Figure 2).[17]

Figure 2. Data use as the main constraint to more effective EMIS across countries

Goal 1: enabling environment	2.4	Goal 2: system soundness	2.1	Goal 3: quality data	2.4	Goal 4: utilization in decision making	1.5
organizational structure	3.2	data architecture	2.5	methodological soundness	2.7	accessibility	2.0
infrastructural capacity	2.7	serviceability	2.5	accuracy and reliability	2.4	disseminating findings	1.8
human resources	2.4	dynamic system	2.4	periodicity and timeliness	2.2	operational use	1.7
legal framework	2.2	data analytics	1.9	integrity	2.0	openness	1.5
budget	2.1	data coverage	1.6				
data-driven culture	2.0						

Note: The figure shows average ratings across 4 goals and 19 dimensions, for the 11 countries that have been surveyed using World Bank SABER-EMIS tool to date. Max score on any dimension is 4.

Concerns with data quality contribute to low utilization.[18] Gathering large quantities of data quickly and for multiple purposes impacts quality. The reliability and accuracy of education administrative data is often questioned, for example, on where teachers are located or how well they are performing.[19] This is for good reason: management decisions such as teacher promotions or the allocation of school grants are high-stakes. But rather than undermining administrative systems, discrepancies highlight the need for better integration of data sources, to improve accuracy.[20] In large-scale system diagnosis and policy analysis, such as in understanding relationships between teacher supply and student attainment, measurement errors for individual observations matter far less. Here imperfect administrative data are well-suited.[21]

More importantly, support to statistical capacity has been driven by global needs and has overlooked country demand.[22] Many international initiatives on statistical capacity building have been excessively supply-driven, focusing on global reporting and accountability pressures, and concerned with improving the availability of indicators or filling gaps in data series.[23] Indeed, a United Nations inspection report covering 43 countries highlights good work on establishing statistical standards but goes on to criticize the lack of support to the effective use of data by national stakeholders.[24] A focus on monitoring global goals has detracted from the local demand for data and subnational analysis needed to actually achieve the goals.[25] As noted in a United Nations Joint Inspection Unit report:

> "If countries are going to achieve the Sustainable Development Goals and go beyond simply monitoring progress towards the targets [then] this requires support for the development of national capacities to explore the wealth of data often produced by national statistical systems and to undertake deeper analysis, not just of trends but of the underlying causes of the obstacles to achieving national development goals."[26]

Severe limits on data linking and access mean that education administrative data are barely used, even by individuals with the relevant skills. The core of most countries' EMIS is the annual school census. This indicates enrolment trends but taken alone it cannot be used to answer, for example, whether school and teacher characteristics—and the policies that affect them—are associated with better learning outcomes,[27] or which schools and districts add value rather than just admitting better students. Internal and external access limitations and a lack of "openness" also contribute to a scarcity of research using national administrative data.[28] In a 2017 review of education data availability across 133 low- and middle-income countries, 61 have no available data and 43 have data only at the national level. Of the 29 countries that do have disaggregated data, the majority are in PDF or non-downloadable format and only 16 countries provide information from student assessments.[29] In a database of papers from 15 leading outlets that publish papers in the economics of education, there are hardly any articles that use administrative data from Africa or Asia over the period 1990–2014 (Table 1). In South America—driven by Chile—there are several entries, but it is in North America and Europe where administrative data are being used to generate context-specific evidence. This is a recent phenomenon, with rapid increase in output over time made possible through enhanced cooperation between governmental agencies and analysts.[30]

Table 1. Number of education studies that make use of administrative data

N. America	Europe	S. America	Africa	Asia	Oceania
253	85	18	6	4	1
		Chile (11)	Kenya (4)	China (3)	Australia (1)
		Colombia (3)	Benin (1)	Taiwan (1)	
		Peru (2)	Uganda (1)		
		Uruguay (1)			
		Venezuela (1)			

Source: Figlio et al. (2015), Education Research and Administrative Data.[31] *Note:* the authors are transparent in stating that "It is possible that this is a substantial overstatement of the North American share of education economics research using administrative data because of possible publication biases of the journals that we selected for inclusion, but again, our charge is not to be comprehensive in our survey."

The underutilization of administrative data matters for educational development. For many public policy questions, findings cannot be imported: the local context is complex and affects both the questions that are most important and the research findings that can answer those questions.[32] Indeed, where we have conflicting estimates of the same intervention across varying contexts, the non-experimental results from the right context are very often a better guide to policy than experimental results from elsewhere.[33] Governments across sub-Saharan Africa and South Asia do make regular investments in collecting data that are relevant to education decision-makers, but these are rarely available and barely used. In a recent study summarising 27,000 economics journal articles written on the 54 countries of Africa, 45 percent are about five countries, accounting for just 16 percent of the continent's population.[34] Where more data are available, countries have higher volumes of research. Elsewhere, underutilisation matters: where access to data and research materials are limited, growth and development are slower.[35] Policymakers concerned about the lack of locally relevant evidence can help by lowering the costs associated with obtaining data for both local and foreign analysts.[36]

Case study 1. Competitively funded education research centers in Chile[37, 38, 39]

Chile

A lesson from Chile's educational reforms is that "none of them would have taken place if it not for a relative abundance of quality data and research."

Main features

- **Opening access** to administrative data through 2000s
- **Research partnerships** with national universities and think tanks
- **Financial support to data use** through a competitive fund for Education Research Centers

Availability of student achievement data in the 1990s attracted a range of international and local researchers. In the mid-2000s the National Science and Technology Agency launched a competitive fund for education research centers. Several centers now exist -- e.g. at the University of Chile -- and the country has a large policy network of experts in universities, think tanks, and foundations. These individuals work with national data and influence policy with their ability to build consensus on technical issues including, most recently, the Bachelet government's reforms that restructured teacher careers. As seen earlier, there are more top-quality papers using Chilean administrative data than are available using similar data from all African and Asian countries combined.

## A SOLUTION: INTEGRATE AND OPEN DATA, THEN SUPPORT ANALYSTS TO USE IT

If we want to see evidence used to drive improvements in education systems, then governments and government data should be our priority. Here, I discuss what may be needed to see the trends in North America—and Chile—discussed above mirrored or surpassed in low- and middle-income countries over the next decade.

1. Integrate datasets

Good, clean EMIS data are useful, but linking EMIS with other administrative datasets— public exams, national assessments, school inspection, teacher payroll—will vastly increase its potential. Efforts to increase EMIS availability, through "data dashboards," for example, may improve accounting and management for activities that are logistical in nature.[40] But ensuring learning and schooling for all is complex, and cannot be achieved without a dramatic change in the way we combine and use data. Linked data allow us to examine things like equity in resource allocation, including key dimensions such as teacher quality, or identify relatively high-performing districts and schools and test relationships between school inspection and student achievement.

Once integrated, administrative data offer opportunities to answer questions of types —and in ways—that are impossible with survey data. The large datasets offer *enormous statistical power*, often with millions of observations in which it is possible to detect modest but meaningful relationships with precision. When used well, they provide the level of disaggregation that officials need for policy applications (Figure 3) and allow the *study of heterogeneous effects* of educational policies: it becomes possible to see how effects vary across different groups of individuals, and if they differ, how they differ, and for whom.[41] Rare events which arrive suddenly, such as disease outbreaks, climatic events, or other policy-related natural experiments, almost certainly do not have prospective data collection, but may be studied in high-frequency administrative data.[42] Furthermore, *working with "real-time" data*

Figure 3. Education ministries need the high-frequency data linked to subnational units that is available in administrative data but less so in nationally representative surveys
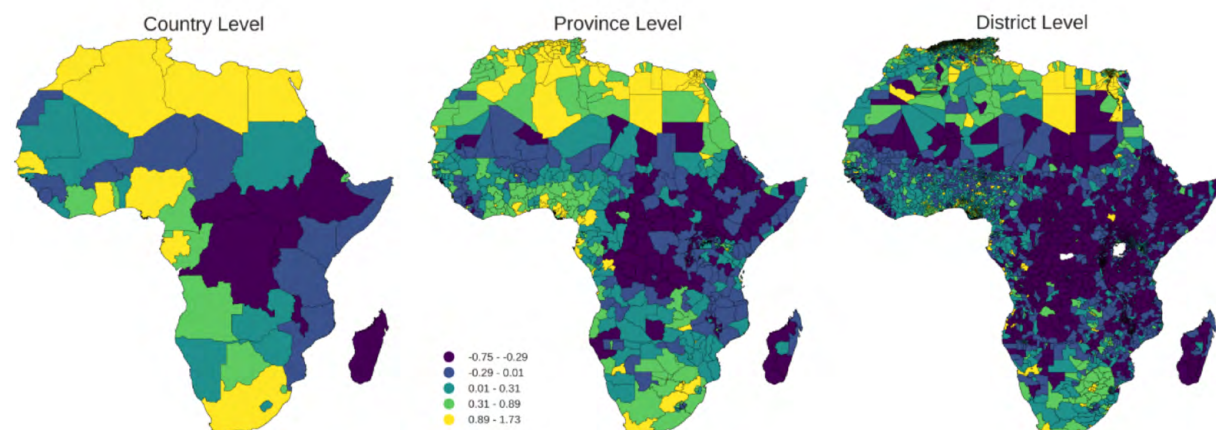
Note: Change from country to province to district level when assessing variation in wealth. Governments require the third figure (and arguably even more disaggregation) for policy applications.

Source: AtlasAI, https://medium.com/atlasai/the-tyranny-of-national-statistics-d3a79af526a9

provides opportunities to study the effects of recent policies and increases the chances of analysis being relevant to current choices and decisions.[43]

Because data coverage is universal, it is feasible to link data from one domain (e.g. education) to data from another domain (e.g. social protection).[44] Making these connections rapidly expands the value of each individual dataset and, for example, knowing whether a district's children are mostly very poor matters in understanding and responding to that district's performance.

Careful data integration can also improve data quality—both its accuracy and its serviceability. A technical perspective on quality refers to completeness, consistency, accuracy: characteristics that make data appropriate for a specific use; but as seen in the Listening to Leaders data, senior government officials want data that are timely, relevant, and actionable.[45] Do available data predict anything about continued enrolment or achievement, or school safety? If not, then why are they being collected? Which indicators are no use? Which are treacherous? The more that countries combine and interrogate education administrative data, the quicker they can identify what is worth collecting, what has limited analytical value, and what is missing (an approach that is being used in Sierra Leone). This is a type of user-centered design, working back from needs and adapting the data systems to satisfy these; it is quite the opposite of traditional supply-side approaches to EMIS design.

Case study 2. Predictive analysis using population-level data in Sierra Leone[46]

Sierra Leone

"Data scientists at Sierra Leone's Directorate of Science, Technology and Innovation are using advanced analytical methods to explore and analyze administrative data collected from schools."

Main features

- **Integrated** enrolment and exams data for all schools
- **Data use partnership** across government departments
- **Public access** to disaggregated data via Education Data Hub

In 2018, the government of Sierra Leone abolished school fees and launched the Free and Quality Education Program. The Ministry of Basic and Senior Secondary Education partnered with the newly formed Directorate of Science, Technology and Innovation to take advantage of their skills. The directorate integrates education administrative data, provides open access and uses predictive analysis, data visualization and machine learning methods to analyze administrative data collected from schools. The partnership makes it easier for ministry officials to understand the statistics, the factors that affect the performance of students and to use the results for resource allocation.

As we have seen, however, investments in the skills for integration and use are rarely made, leaving data units in most ministries understaffed. One of the lessons of the past 20 years is that top-down initiatives have not led to sustained increases in capacity.[47] In the education sector, while projects have for decades sought to develop human capital, skills development has typically focused on education planning, with little attention paid to developing programs for statistics or IT skills in the education sector.[48] Statistical Capacity Indicators show that overall capacity in government is lowest in sub-Saharan Africa and although several countries have made progress, several others have fallen back (Figure 4); there has been very little improvement overall in the 15 years to 2018.

Figure 4. Overall statistical capacity has increased (slightly) in sub-Saharan Africa—and more so in South Asia—from 2004 to 2018
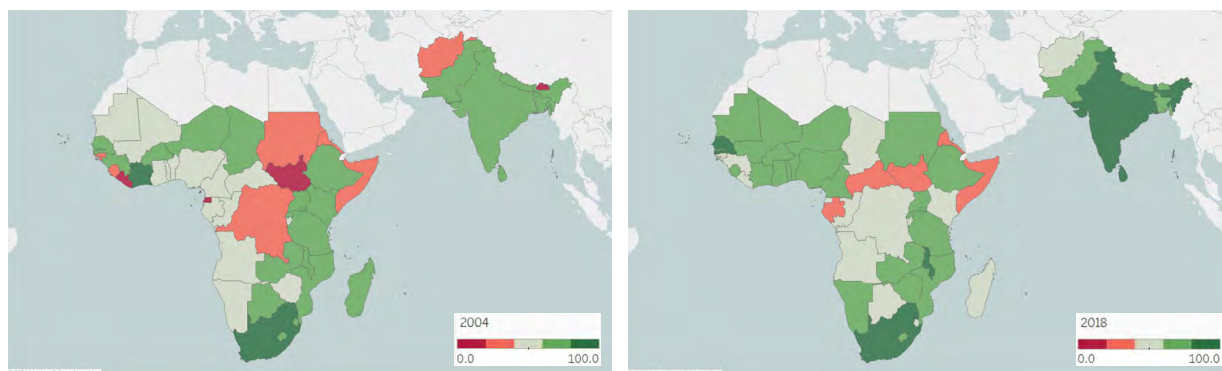
Source: World Bank Statistical Capacity Indicators Database, mapped for 2004 (earliest year) and 2018 (most recent year).

2. Make data open access so officials, social scientists, and practitioners can contribute skills and generate insights

Linking datasets is important, but the step change comes from providing access for motivated individuals and organizations to work with administrative data. Persistently low internal capacity does not mean that progress must remain slow. There are individuals with the skills, incentives, and time, who may not work for government but who are willing and able to collaborate to solve specific data problems. Opening access is a strategy to draw in these specialists to facilitate system diagnosis and policy evaluation.[49]

A major strength of administrative data is its (potentially) immediate availability, at relatively low additional cost. Partner institutions can be enlisted to support data linking, hosting and sharing, removing that burden from ministries of education (e.g. in the case of South Africa), and these data can be used across government offices for management and operational decisions. Rather than supplanting them—as is often the case with surveys or large-scale international assessments—public-use administrative data platforms increase the analytical value of existing national systems.

To get there, the gains in performance earned through better analysis must outweigh the perceived losses from public oversight. Incentives to increase data access are not always aligned within governments. Governments need data and insights to make better decisions and officials may even find that increasing data availability is a way of signalling a commitment to certain reforms.[50] But these same data may expose the shortcomings of policies and lead to demands for change.[51] Governments may fear that the data may be used to present too simplistic a picture in a complex system, or that their release may hamper public support for policies that will work, if only "given enough time."[52] For politicians facing electoral pressures, releasing data may undermine the credibility of campaign promises or provide ammunition to the opposition.

Case study 3. Efficiency gains through data partnerships in South Africa[53]

South Africa

"Researchers who access data also are required to present the research output to the department. This is to ensure that greater access to data actually leads to better informed policy and practice in the South African government."

Main features

- **Data hosted by third-party** providing management and documentation benefits to government
- **Simple protocols** ensure privacy is maintained and laws are met
- **Impact evaluations of national policies** have been possible because of access to these administrative data

In the South African Department of Basic Education data use is being extended in at least two ways. (1) a set of protocols has been developed to ensure a transparent application process for external researchers to request data, to review requests and to facilitate the sharing of data. Researchers are also required to present findings to officials, ensuring that greater data access leads to better informed policy and practice. (2) a partnership has been established with DataFirst to share administrative data through an online portal -- one of numerous examples showing that it is possible to provide secure access to de-identified administrative data. Thanks to these initiatives, at least three impact evaluations (of the Grade R Preschool Programme, the Language in Education Policy, and the impact of sub-national administrations) have been possible.

There are practical reasons why officials are hesitant to share the data that they are entrusted to protect.[54] A challenge in navigating data access stems from the absence of legal frameworks and clear responsibilities in many systems.[55] The production and dissemination of official statistics is the core or even exclusive task of national statistical offices but for other producers, such as ministries of education and exam bodies, generating data and statistics is one of several tasks, not the most important.[56] Engaging statistics offices can often help to avoid conflicts of interest and improve connections between producers and users.[57, 58] Finding an appropriate solution in each setting is not straightforward, but there is precedent (including the cases in this note) for finding practical ways of opening access. These need not require enormous changes to practice: clear data policies can afford, for example, (a) direct access to microdata stripped of individual identifying information, (b) fair and transparent processes for access, and (c) systems that ensure immediate disclosure of statistical results and findings to interested government stakeholders.

The major benefit of increasing access is the opportunity for officials, social scientists, and practitioners to combine skills and generate better ideas. Getting data into the hands of the right individuals and institutions can support government systems with advanced skills and quickly change the way that insights are produced, with large spillover benefits.[59] Instead of waiting until ministry data units have all the expertise, by working across institutions, governments and their staff benefit from improved approaches to data analysis, providing a faster path to evidence and one which offers substantial benefits for skills development.[60] It isn't just a matter of presenting data in more impressive formats;[61] these data can tell us what small experiments miss: general equilibrium effects, political economy, and government implementation challenges.

By opening access governments can leverage their data and encourage analysts to provide answers to specific questions. In the right hands, even imperfect administrative data are well suited to system diagnosis and policy evaluation, particularly where issues are stark (i.e. where we are not trying to identify tiny margins) and persistent (i.e. the ones that officials probably want to tackle the most).

Statistical capacity within government is not improving quickly, but a supply of integrated government data solves a problem for analysts who can be incentivized to tackle questions that governments want answers to (an approach that has been used in Peru). Another approach would be to bring in analytical capacity to support the process, but this would reduce the efficiency and skills benefits from working across institutions.

Case study 4. Leveraging government data to attract analysts to Peru[62]

Peru

"MineduLAB intentionally connects Peruvian researchers to opportunities to conduct research with the Ministry of Education's administrative data."

Main features

- **Government leverages data** to attract skills from international research organisations
- **Emphasis on cost-effectiveness** of new policies, which may then be taken to scale
- **Evaluations use administrative data**: so lower-cost and quicker than with primary data collection
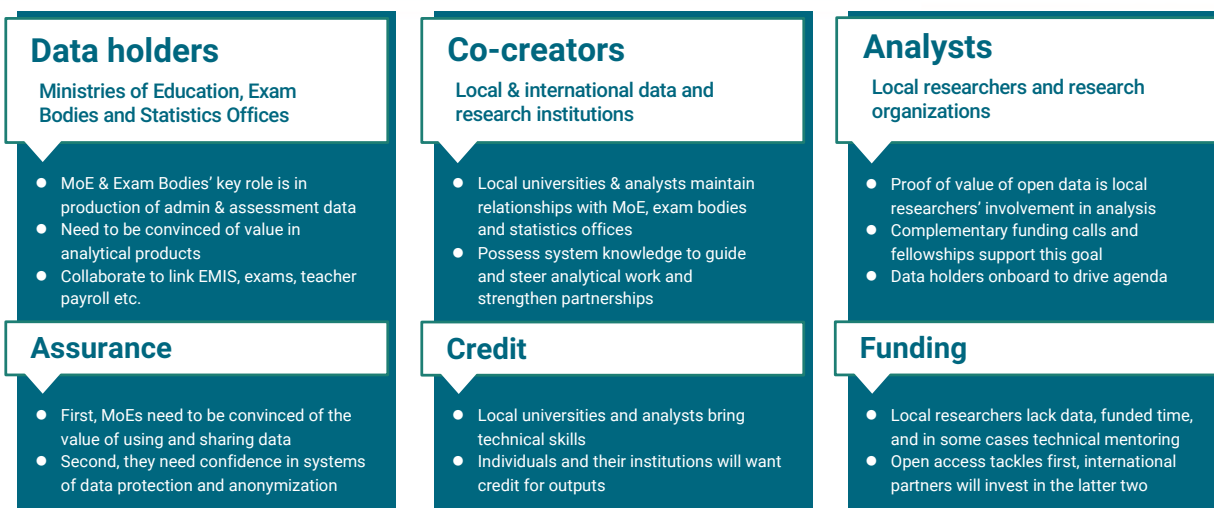
MineduLAB was developed as a collaboration between the Ministry of Education in Peru, J-PAL, IPA and the World Bank. MineduLAB connects researchers to opportunities to conduct research with the Ministry of Education's administrative data. Working with researchers in this way contributes to the development of a wider field of actors promoting data and evidence use in policy-making, by influencing and connecting their academic institutions and government network. MineduLAB prioritises innovations that can be implemented at low cost and evaluated using existing administrative data, which reduces barriers to piloting and scaling up. Good administrative data significantly reduce the costs of evaluation, including at scale, making it easier to generate evidence on ongoing policy commitments.

HOW TO MAKE IT HAPPEN

Integrating administrative data and making it accessible to officials, social scientists, and practitioners would capitalize on the huge investments made in data and allow governments to fast-track the generation of ideas. This note focuses on national administrative systems because they are comprehensive, representative, contextually relevant, and can be used to meet local demand for insights on management, operational and policy decisions:

- *Informing management:* which teachers are licensed; how much are they being paid and does pay arrive on time; how has the number of girls enrolled in a district changed in the past three years; what are the warning signs of dropout?

- *Improving operational processes and choices:* which school inspection indicators best predict student learning; at what ages do students leave the system and where should additional support be provided; what is the sampling frame—with contextual variables—for a national learning assessment?

- *Guiding policy reform:* what is the relationship between language of instruction and student performance; what are the characteristics of the districts that are adding most value to student achievement, relative to resources; and what was the impact, across provinces, of the teacher salary reform on girls' attainment?

Figure 5. Three layers of partnerships required to make it happen

Data holders	**Co-creators**	**Analysts**
Ministries of Education, Exam Bodies and Statistics Offices	Local & international data and research institutions	Local researchers and research organizations
• MoE & Exam Bodies' key role is in production of admin & assessment data • Need to be convinced of value in analytical products • Collaborate to link EMIS, exams, teacher payroll etc.	• Local universities & analysts maintain relationships with MoE, exam bodies and statistics offices • Possess system knowledge to guide and steer analytical work and strengthen partnerships	• Proof of value of open data is local researchers' involvement in analysis • Complementary funding calls and fellowships support this goal • Data holders onboard to drive agenda

Assurance	**Credit**	**Funding**
• First, MoEs need to be convinced of the value of using and sharing data • Second, they need confidence in systems of data protection and anonymization	• Local universities and analysts bring technical skills • Individuals and their institutions will want credit for outputs	• Local researchers lack data, funded time, and in some cases technical mentoring • Open access tackles first, international partners will invest in the latter two

Source: Adapted from a CGD Education Programme internal note prepared by Justin Sandefur.

Partnerships are required to make this happen. Progress requires a step away from traditional approaches of strengthening data units within ministries. Several of the countries that are making the most progress in drawing analytical value from administrative data have done so through increasing access and developing strong partnerships with universities and other agencies, where most of the analytical work is done. Partnerships at three levels will change who is involved in the processing and use of data and how they interact (Figure 5).

Success rests on decision-makers seeing the value in using and sharing data. First, by enlisting co-creators, governments benefit from specialist skills in data linking and can rapidly increase the pace of integration and sharing. Data and research institutions holding relationships with ministries of education, exam bodies, and statistical offices bring contextual knowledge and can provide assurances regarding systems of data protection and anonymization. Second, together with education ministries, co-creators can steer an analytical agenda for local research organizations. The supply of integrated data solves a problem for these organizations which can dedicate the time to producing targeted outputs. Third, individuals in these organizations already have the skills and incentives to work with data but their engagement will usually require financial, and in some cases technical, support. These resources may be provided by government, e.g. as part of a call for papers, or by nongovernmental organizations.

Failure would be signalled by an inability of data users to produce actionable ideas, or by ideas having no impact on educational development. The associations between administrative data availability and the production of ideas,[63] and between ideas and development, are supported by a growing body of literature.[64] But it is not certain that they will hold across low- and middle-income countries, or in any education sector with limited capacity or initial interest in using data.[65] Should these ideas find traction then it may be possible to contribute to the evidence on connections. Equally, access to more and better data cannot always be the solution and these signals—i.e. that data access is not increasing analytical output, or that this output is not entering the policy-making process—may serve as useful stopping conditions.

This proposal is pragmatic. Scarce skills are needed to integrate and analyze data, and these are not always within government. Data cleaning, linking, archiving and documentation is tedious work and it is hard to focus efforts on these tasks, making it difficult to get the organization to go that last mile, even if it is committed to it in principle.[66] The proposed approach requires collaboration to draw in skills and expertise of those incentivized to complete the job, adding strength to the overall statistical system. Cooperation of this type will arise if all parties can share the benefits of the output and precedents for this kind of cooperation exist, including those country cases described in this note.

It can also rapidly improve the use and usefulness of national administrative data. By working backwards from the bureaucrats', policymakers', and school leaders' demands for information, partners can work with integrated data to quickly identify what is worth collecting, what has limited analytical value and what is missing. This provides a far faster path to relevant data and insights than a redesign in a central office, based on general concerns with data coverage in questionnaires and reliability of leading indicators.

By prioritizing integration and open access, it is possible to fundamentally shift the way that education data *are used* to generate ideas. The proposed approach is not interested in using data to create a top-down, command and control education system; it is about linking and then using administrative data to generate ideas for educational development. Investments in integrating national data are relatively cheap and offer insight into a wide range of questions. Providing a high degree of access to these data makes empirical analysis in education more feasible, more believable and more relevant. The whole purpose of any data system is for its output to be used—this approach can rapidly increase how existing data are used to produce good ideas.[67]

ENDNOTES

1   See e.g. Das, J., Do, Q.T., Shaines, K., Srikant, S., 2013. US and them: the geography of academic research. J. Dev. Econ. 105, 112–130; Furman, J., Stern, S., 2011. Climbing atop the Shoulders of Giants: The Impact of Institutions on Cumulative Research. American Economic Review, 101 (5): 1933–63.

2   Pritchett, L., Sandefur, J., 2014. Context Matters for Size: Why External Validity Claims and Development Practice do not Mix, Journal of Globalization and Development, De Gruyter, vol. 4(2), 161–197.

3   Card, David, Raj Chetty, Martin Feldstein, and Emmanuel Saez. 2010. Expanding access to administrative data for research in the United States. Washington, DC: National Science Foun- dation White Paper No. 10–069.

4   UNESCO Institute for Statistics (2017), The Data Revolution in Education, Information Paper No. 39. Paris: UNESCO; & Read, L,. Atinc, T., 2017. Information for Accountability: Transparency and Citizen Engagement for Improved Service Delivery in Education Systems, Brookings Global Economy and Development Working Paper 99.

5   Custer, S. Sethi, T. (Eds.), 2017. Avoiding Data Graveyards: Insights from Data Producers and Users in Three Countries. Williamsburg, VA: AidData at the College of William & Mary.

6   77 countries that qualified for concessional borrowing through the International Development Association (IDA) at the time of the original report.

7   Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development (2015). Sustainable Development Solutions Network.

8   Custer, S., King, E., Atinc, T., Read, L., Sethi, T., 2018. Toward Data-Driven Education Systems: More Data and More Evidence Use, Journal of International Cooperation in Education, Vol.20–2/21–2 (2018) pp. 5–34

9   PARIS21 (2019), "Partner Report on Support to Statistics 2019", Paris.

10 Honig, D., Pritchett, L. (2019). The Limits of Accounting-Based Accountability in Education (and Far Beyond): Why More Accounting Will Rarely Solve Accountability Problems, CGD Working Paper 510.

11 Abdul-Hamid, Husein, Namrata Saraogi, and Sarah Mintz. 2017. Lessons Learned from World Bank Education Management Information System Operations: Portfolio Review, 1998–2014. World Bank Studies. Washington, DC: World Bank. doi:10.1596/978-1-4648-1056-5.

12 Abdul-Hamid, Husein, Namrata Saraogi, and Sarah Mintz. 2017. Lessons Learned from World Bank Education Management Information System Operations: Portfolio Review, 1998–2014. World Bank Studies. Washington, DC: World Bank. doi:10.1596/978-1-4648-1056-5.

13 Abdul-Hamid, Husein, Namrata Saraogi, and Sarah Mintz. 2017. Lessons Learned from World Bank Education Management Information System Operations: Portfolio Review, 1998–2014. World Bank Studies. Washington, DC: World Bank. doi:10.1596/978-1-4648-1056-5.

14 Aiyar, Y., Bhattacharya, S. (2016). The Post Office Paradox, A Case Study of the Block Level Education Bureaucracy. Economic & Political Weekly, Vol. 11, No. 11.

15 Abdul-Hamid, Husein, Namrata Saraogi, and Sarah Mintz. 2017. Lessons Learned from World Bank Education Management Information System Operations: Portfolio Review, 1998–2014. World Bank Studies. Washington, DC: World Bank. doi:10.1596/978-1-4648-1056-5.

16 World Bank (2017). Program Appraisal Document on a Proposed IDA Grant and Multi-Donor Trust Fund Grant to the Federal Democratic Republic of Ethiopia for the General Education Quality Improvement Program for Equity (GEQIP-E).

17 For full information on the World Bank's SABER program, including methods and terminology please see: http://saber.world-bank.org/index.cfm

18   Abdul-Hamid, Husein, Namrata Saraogi, and Sarah Mintz. 2017. Lessons Learned from World Bank Education Management Information System Operations: Portfolio Review, 1998–2014. World Bank Studies. Washington, DC: World Bank. doi:10.1596/978-1-4648-1056-5.

19   Asim, S., Chimombo, J., Chugunov, D., Gera, R., 2017. Moving Teachers to Malawi's Remote Communities A Data-Driven Approach to Teacher Deployment. Policy Research Working Paper 8253, Washington DC: World Bank.

20   Custer, S. Sethi, T. (Eds.), 2017. Avoiding Data Graveyards: Insights from Data Producers and Users in Three Countries. Williamsburg, VA: AidData at the College of William & Mary.

21   Figlio, David N.; Karbownik, Krzysztof; Salvanes, Kjell G. (2015). Education Research and Administrative Data, IZA Discussion Papers, No. 9474, Institute for the Study of Labor (IZA), Bonn

22   OECD (2017), Development Co-operation Report 2017: Data for Development, OECD Publishing, Paris. http://dx.doi.org/10.1787/dcr-2017-en

23   Open Data Watch (2015). Navigating the politics of open data: A High-Level Political Forum (HLPF) Side Event. Outcome Document.

24   United Nations Joint Inspection Unit (2016). Evaluation of the contribution of the United Nations development system to strengthening national capacities for statistical analysis and data collection to support the achievement of the Millennium Development Goals and other internationally agreed development goals. A/71/431, JIU/REP/2016/5

25   Demombynes, G., Sandefur, J., 2014. Costing a Data Revolution, Data for Development Perspective Paper, Copenhagen Consensus Center.

26   United Nations Joint Inspection Unit (2016). Evaluation of the contribution of the United Nations development system to strengthening national capacities for statistical analysis and data collection to support the achievement of the Millennium Development Goals and other internationally agreed development goals. A/71/431, JIU/REP/2016/5

27   Lucas, Adrienne M., and Isaac M. Mbiti. 2012. "Access, Sorting, and Achievement: The Short-Run Effects of Free Primary Education in Kenya." American Economic Journal: Applied Economics, 4 (4): 226–53.

28   Figlio, David N.; Karbownik, Krzysztof; Salvanes, Kjell G. (2015). Education Research and Administrative Data, IZA Discussion Papers, No. 9474, Institute for the Study of Labor (IZA), Bonn

29   Brookings information for accountability. READ https://www.brookings.edu/wp-content/uploads/2017/01/global_20170125_information_for_accountability.pdf

30   Figlio, D., Karbownik, K., Salvanes, K., 2016. Education Research and Administrative Data, in Hanushek E., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education, Vol. 5.

31   This table lists studies which make use of administrative data and were published between January 1990 and July 2014 in the following outlets: American Economic Journal: Applied Economics, American Economic Journal: Economic Policy, American Economic Review, Econometrica, Economic Journal, Economics of Education Review, Education Finance and Policy, Journal of Human Resources, Journal of Labor Economics, Journal of Political Economy, Journal of Public Economics, Quarterly Journal of Economics, Review of Economic Studies, Review of Economics and Statistics and NBER Working Papers: Education Program.

32   Nunn, N., 2019. Rethinking Economic Development, Canadian Journal of Economics, Vol. 52, Issue 4, pp. 1349–1373.

33   Pritchett, L., Sandefur, J., 2014. Context Matters for Size: Why External Validity Claims and Development Practice do not Mix, Journal of Globalization and Development, De Gruyter, vol. 4(2), 161–197.

34   Porteous, O., 2020. Research Deserts and Oases: Evidence from 27 Thousand Economics Journal Articles on Africa.

35   Das, J., Do, Q.T., Shaines, K., Srikant, S., 2013. US and them: the geography of academic research. J. Dev. Econ. 105, 112–130.

36   Porteous, O., 2020. Research Deserts and Oases: Evidence from 27 Thousand Economics Journal Articles on Africa.

37 Emiliana Vegas: 5 lessons from recent educational reforms in Chile, Brookings: https://www.brookings.edu/research/5-lessons-from-recent-educational-reforms-in-chile/

38 Mizala, A., Schneider, B., 2019. Promoting quality education in Chile: the politics of reforming teacher careers, Journal of Education Policy, DOI: 10.1080/02680939.2019.1585577

39 See database introduced in Table 1.

40 Honig, D., Pritchett, L. (2019). The Limits of Accounting-Based Accountability in Education (and Far Beyond): Why More Accounting Will Rarely Solve Accountability Problems, CGD Working Paper 510.

41 Roed, K., Raaum, O., 2003. Administrative Registers: Unexplored Reservoirs of Scientific Knowledge? Econ. J. 113 (488), F258–F281.

42 Card, David, Raj Chetty, Martin Feldstein, and Emmanuel Saez. 2010. Expanding access to administrative data for research in the United States. Washington, DC: National Science Foun- dation White Paper No. 10–069.

43 Einav, L., Levin, J., 2014. The Data Revolution and Economic Analysis, Innovation Policy and the Economy, 14, issue 1, p. 1–24.

44 Figlio et al. (2017). The Promise of Administrative Data in Education Research. Presidential Essay, Association for Education Finance and Policy.

45 Custer, S., King, E., Atinc, T., Read, L., Sethi, T., 2018. Toward Data-Driven Education Systems: More Data and More Evidence Use, Journal of International Cooperation in Education, Vol.20–2/21–2 (2018) pp. 5–34

46 Assaitou Bah, Department of Science, Technology and Innovation, Sierra Leone, direct communication.

47 Kiregyera (2013). The emerging data revolution in Africa: Strengthening the statistics, policy and decision-making chain. Sun Press.

48 Crouch, L., 2019. Meeting the Data Challenge in Education, A Knowledge and Innovation Exchange (KIX) Discussion Paper. Washington DC: Global Partnership for Education.

49 https://esrc.ukri.org/news-events-and-publications/news/news-items/addressing-major-societal-challenges-by-harnessing-government-data/

50 Taylor, M. (2016). The Political Economy of Statistical Capacity, A Theoretical Approach. Discussion Paper No. IDB-DP-471, Inter-American Development Bank.

51 Taylor, M. (2016). The Political Economy of Statistical Capacity, A Theoretical Approach. Discussion Paper No. IDB-DP-471, Inter-American Development Bank.

52 Taylor, M. (2016). The Political Economy of Statistical Capacity, A Theoretical Approach. Discussion Paper No. IDB-DP-471, Inter-American Development Bank.

53 Stephen Taylor, Department of Basic Education, South Africa, direct communication.

54 Figlio et al. (2017). The Promise of Administrative Data in Education Research. Presidential Essay, Association for Education Finance and Policy.

55 Open Data Watch (2019). Navigating the politics of open data: A High-Level Political Forum (HLPF) Side Event.

56 United Nations Economic Commission for Europe (2008). How Should a Modern National System of Official Statistics Look? UNECE, Statistical Division.

57 United Nations Economic Commission for Europe (2008). How Should a Modern National System of Official Statistics Look? UNECE, Statistical Division.

58   NSOs are also respected for maintaining stable statistical concepts—which can be crucial for policy applications.

59   Duflo, E. 2017. The Economist as Plumber. Essay based on the Ely lecture, AEA meeting 2017.

60   Punton, M., 2016. How Can Capacity Development Promote Evidence-Informed Policy Making, Literature Review for the Building Capacity to Use Research Evidence (BCURE) Programme. ITAD.

61   Crouch, L., 2019. Meeting the Data Challenge in Education, A Knowledge and Innovation Exchange (KIX) Discussion Paper. Washington DC: Global Partnership for Education.

62   Creating a Culture of Evidence Use: Lessons from J-PAL's Government Partnerships in Latin America (2018). https://www.povertyactionlab.org/sites/default/files/documents/creating-a-culture-of-evidence-use-lessons-from-jpal-govt-partnerships-in-latin-america_english.pdf

63   Currie, J., Kleven, H., Zwiers, E. (2019) Technology and Big Data Are Changing Economics: Mining Text to Track Methods. Prepared for AEA/ASSA Session: Empirical Practice in Economics: Challenges and Opportunities.

64   Das, J., Do, Q.T., Shaines, K., Srikant, S., 2013. US and them: the geography of academic research. J. Dev. Econ. 105, 112–130.

65   Figlio, D., Karbownik, K., Salvanes, K., 2016. Education Research and Administrative Data, in Hanushek E., Machin, S., Woessmann, L. (Eds.), Handbook of the Economics of Education, Vol. 5.

66   Singh, A., via Opening up Microdata Access in Africa, http://blogs.worldbank.org/impactevaluations/opening-up-microdata-access-in-africa

67   Figlio, David N.; Karbownik, Krzysztof; Salvanes, Kjell G. (2015). Education Research and Administrative Data, IZA Discussion Papers, No. 9474, Institute for the Study of Labor (IZA), Bonn

Center for Global Development

Ideas to Action: Independent research for global prosperity

JACK ROSSITER is a senior policy analyst at the Center for Global Development

WWW.CGDEV.ORG