

To take learning seriously, start by measuring it



“The results tell us about the fact of the drop in education quality, after releasing the 2015 results of the Trends in International Mathematics and Science Study. . . . We cannot ignore what is happening to our education, and we cannot afford the repercussions of not reforming it.”

QUEEN RANIA OF JORDAN, FACEBOOK POST, DECEMBER 2016

Why does the learning crisis persist? How can children attend school for years but remain functionally illiterate? Why don't the people in education systems fix this? One big reason is that, for many, the learning crisis is invisible. Education systems have little systematic information on who is learning and who is not. As a result, it is impossible to generate an impetus for action—let alone a plan.

To tackle the crisis, it is necessary—though not enough—to measure learning. But learning metrics must facilitate action, be adapted to country needs, and consist of a range of tools to meet the needs of the system, including at the classroom level.

The learning crisis is often hidden—but measurement makes it visible

“Almost no low-income countries have standardized (equated over time) national assessment systems to track learning and provide a feedback mechanism to national education policies and programs” (Birdsall, Bruns, and Madan 2016, 2).

Education systems routinely report on enrollment—but not on learning. Because learning is missing from

official education management data, it is missing from the agendas of politicians and bureaucrats. This is evident in how politicians often talk about education only in terms of inputs—number of schools, number of teachers, teacher salaries, school grants—but rarely in terms of actual learning. Lack of data on learning means that governments can ignore or obscure the poor quality of education, especially for disadvantaged groups.

Without objective information on learning, parents may be unaware of the poor quality of education. This prevents them from demanding better services from schools and governments. In Kenya, one study found that less than half of the children in grade 4 could pass basic proficiency tests in literacy or numeracy, yet more than two-thirds of adults were broadly satisfied with the government's performance in education.¹ The realization that learning outcomes are poor may come only when children face poor labor market prospects, but by then it is too late. If parents have no real information on how much (or little) their children are learning, how can they hold schools or governments accountable?

Without clear information on what students do not know, how can schools improve instruction? Teachers may find it hard to judge to what extent students understand what is being taught. This is

particularly true in low-income countries, where teachers face large classrooms that mix students of very different abilities. For example, a study from Delhi, India, found that the same grade may contain students whose achievement level spans the equivalent of five to six grades.² In such contexts, learning measures provide teachers with timely feedback about which students may need additional support. More broadly, these measures provide school management with information about which areas need attention to improve instruction. If the information is shared with parents or students, it can help them direct their own efforts toward improving learning.

Yet concerted action is often derailed by concerns about the possible pitfalls of learning metrics. These metrics generate much debate on, for example, the outsize impacts of international assessments on local policy, the limited use of national assessments for improving classroom practice, or the potential gaming of high-stakes testing.³ But *measurement of learning* is not shorthand for international testing such as the Programme for International Student Assessment (PISA) or for the high-stakes accountability approach implemented through the U.S. No Child Left Behind policy. Instead, the term covers a range of assessments, including formative classroom assessments (box 4.1). Even in this form, measures of learning provide information on only some of the skills students acquire as they develop (see spotlight 3 on multidimensionality of skills). Thus metrics are

complements of, not substitutes for, careful, context-specific analysis to determine how to improve learning.⁴

Measures for learning guide action

Testing in Rio de Janeiro, Brazil, happens at two levels. First, every two years all students in grades 5 and 9 take a national test (the Prova Brasil) designed to assess public education. Second, students are tested at the end of each two-month curriculum block. These tests, given by municipal education departments, aim to provide quick feedback to teachers and principals, allowing schools and the broader system to provide more support to struggling students (Elwick and McAleavy 2015).

Identifying learning gaps in the classroom is the first step toward resolving them. In environments of low learning, there is often a gap between the level of students and the level at which classes are being taught.⁵ This might be because teachers are unaware of students' levels. Fostering a culture of classroom-based assessments can address this problem. In Singapore, students are given screening tests at the start of grade 1, which helps teachers identify those who require additional instruction to learn to read.⁶

Learning metrics help highlight where support is most needed. School districts and schools are then

Box 4.1 Good measures of learning illuminate all parts of the education system

Formative classroom assessments facilitate instruction by providing real-time feedback to support teaching and learning. This feedback allows teachers to identify struggling students, thereby enabling them to adjust instruction to meet the learning needs of different students. Classroom assessments also generate valuable feedback for students and parents.

National assessments provide information on the overall education system by highlighting achievements along with challenges, such as inequalities. They are useful for education management, policy, and reform.

National examinations certify student achievement, with a focus on transparently selecting students for more

advanced placements in the education system or job market. Because of their role in determining labor market outcomes, these examinations are high-stakes for students. They significantly affect what is taught and how, and they are critical for managing the flow of students through the system.

International assessments benchmark student performance by evaluating education systems across countries and over time using representative samples of children. There has also been a steady increase in the use of citizen-led assessments. These can be important for fostering public awareness, showing what is possible, advocating for change, and informing research.

Source: WDR 2018 team.

better able to target resources to improve service delivery. In Brazil, national assessments have been widely adopted by states and municipalities to strengthen school performance.⁷ Learning metrics have also guided big-banner education reforms. In Chile, PISA's reading framework guided national curriculum reform.⁸ Similarly, findings from the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) I, 1995–99, underpinned a review of Mauritius's education master plan.⁹ In some cases, learning metrics have been instrumental in making education reform data-driven. In Germany, lower than expected results—especially for students from poorer backgrounds—on the 2000 PISA led to the development of more support for disadvantaged students, especially those from immigrant backgrounds.¹⁰

For learning metrics to guide action effectively, they need to be used as a range of tools to serve different needs, from classroom practice to system management. Measures of learning come in various forms, with different measures serving different purposes for different actors. These range from simple oral questions posed by a teacher to national assessments that help policy makers prioritize action (box 4.1). In well-functioning systems, these different tools complement one another to form a coherent whole.¹¹

Policy makers should rely on a broad range of information instead of any one measure. When a single metric becomes the sole basis for big policy triggers, the corresponding stakes may become dangerously high. A striking example is the U.S. No Child Left Behind policy enacted in 2001. This policy had strong negative repercussions for schools that performed poorly on annual statewide standardized tests. Though the policy led some poorly performing schools to improve, it also generated various undesirable strategic responses by teachers and school administrators.¹² These included reclassifying students as requiring special education, exempting certain students from testing, reallocating resources to students at the margin of passing, and suspending low-scoring students near test dates.¹³ Even in the case of PISA, some studies have suggested that the performance in some places—Argentina, Malaysia, Vietnam, Shanghai (China)—could be tied in part to (perhaps inadvertent) “selective samples” that may exclude some poorly performing schools or students.¹⁴

Education systems also routinely underuse the information generated by learning metrics—making for a lot of measurement that leads to little action.¹⁵ Often, findings are simply not communicated in a timely way to relevant audiences.¹⁶ There may also be credibility issues. If teachers or schools do not feel

heard or acknowledged in a national assessment process, they will likely reject its findings. For example, teachers are more likely to resist quantitative forms of evaluation when metrics do not take into account context.¹⁷ This is particularly the case for measures of learning disseminated as rankings, which are susceptible to being taken out of context. In some education systems, such friction is heightened by the use of technology, which raises questions about privacy and transparency. Approaches using technology also involve limited social interaction, which is associated with less impact.¹⁸

For measurement to guide action, it must be actionable. It also needs to be available to stakeholders. At the design stage, stakeholders have to ask themselves how learning data will be used. In Chile, all students in grades 4 and 8 take the Sistema de Medición de la Calidad de la Educación (SIMCE) each year. After the test identifies the 900 schools scoring in the lowest 10 percent on the tests in their province, these schools receive special resources. The data, then, are clearly linked to action. Many assessment systems measure outcomes too infrequently or too broadly to be of practical use. The most recent publicly available data from the SACMEQ are for 2007. Another constraint is the lag between when data are collected and when they are made available, as well as how data are made available. Many ministries produce only hard copies of summary reports, which make them difficult to use.

Measures of learning spur action

“Shock as 60 [Percent] of Tanzania Students Fail National Exam” (East African, 2013)

In the United States since 2001, information on different schools' performance on standardized tests has notably increased turnout in local school board elections (Holbein 2016).

Measures of learning motivate action through three channels:¹⁹

- **Participation.** Learning outcomes are often far worse than stakeholders realize. In Uganda, nearly three-quarters of parents said they were satisfied with the quality of education—yet only a quarter of grade 4 students could pass a math test based on grade 2 questions.²⁰ By documenting service delivery shortfalls, learning metrics can motivate parents to hold their schools accountable for learning.

In such contexts, learning metrics can correct information failures, which are especially severe for the poor. This correction can in turn rebalance the relationship between users and providers. This channel operates via the direct or short route of accountability running from parents directly to schools.

- *Choice.* Providing parents with hard evidence about learning outcomes at alternative schools can encourage schools to improve learning by increasing competitive pressures. When parents have objective information about learning outcomes across schools, they can punish poorly performing schools by “voting with their feet.” Public schools care about such outcomes because their resources are often tied to the number of students they enroll.²¹ But this channel may also disproportionately penalize schools that serve poor children.
- *Voice.* Learning metrics can facilitate lobbying for reform by providing information on what needs fixing. Lack of reliable metrics, by contrast, undermines accountability for results.²² This channel operates via the long route of accountability, where learning metrics may help citizens use the political process to hold politicians accountable for learning.

That said, the links from measurement to action are neither automatic nor straightforward. India’s citizen-led assessment, the Annual Status of Education Report (ASER), has documented low proficiency scores since it was introduced in 2004. However, clear or sustained improvements are not yet visible for the country as a whole.²³ At the same time, some Indian states have shown significant improvements in grade 3 reading levels between ASER 2010 and 2016.²⁴ This shows that it is not just the information but action that matters. For learning to improve, not only do learning assessments need to be available, but also someone needs to act on them. In fact, an evaluation of the impact of citizen-led assessments in Kenya finds that for information on learning to spur action, those who receive the information must understand it, see it as actionable, care about the topic, and believe that their actions will improve outcomes.²⁵

Political pressures may limit the extent to which measures of learning spur positive action. Where education quality is low, politicians have an incentive to hide or obscure learning outcomes.²⁶ They may also try to evade blame for poor performance by setting low standards, trying to limit year-to-year comparability, or restricting access to outcome information.²⁷ For example, Argentina amended its standardized test so that year-to-year comparisons are not possible,

decreased the frequency with which the test is administered, and delayed the publication of results by two years to obscure the poor performance of students.²⁸ Teachers, too, might resist learning assessments to minimize opportunities for blame.²⁹ In Chile, teacher training institutions have shown resistance to the national assessment.³⁰ Assessments are also political because they can affect the flow of resources or prestige in an education system—as in the United States under the No Child Left Behind policy.³¹ Underlying politics can make student assessment systems particularly hard to reform (see part IV of this Report).

When does measurement mobilize citizens to demand accountability for learning? Because of limited attention, information is often ignored, especially if it is complex or provides unwelcome news.³² Therefore, for measurement to spur action, information must be available in an easily digestible way. But this in itself may not be enough. Learning metrics can galvanize communities to hold their schools accountable for learning only when collective action problems are resolved.³³ A participatory approach—where schools and communities have a say in what type of “learning metrics” are generated at the school level—may be likely to work better here.³⁴ In addition, for citizens to be able to act on information, fear of reprisals must be low. Finally, for citizens to act in behalf of change, they must believe that their own individual actions can make a difference.³⁵

Efforts to benchmark country performance through international or regional assessments have in some cases galvanized action because international comparisons make learning politically salient. Release of the Trends in International Mathematics and Science Study (TIMSS) or PISA rankings often triggers intense media interest, inserting learning into political and economic debates.³⁶ This increase in interest often generates momentum for government action—an effect known as “PISA shock”—thereby unleashing targeted reforms. About half the countries participating in the PISA assessments under the aegis of the Organisation for Economic Co-operation and Development (OECD) have launched reforms because of the results.³⁷ Learning assessments also spur action by making learning a tangible goal. Whereas the United Nations’ Millennium Development Goals (MDGs), which inspired efforts by governments and donors, focused on enrollment, the current Sustainable Development Goals (SDGs) place greater emphasis on learning.³⁸ The success of the SDGs will depend on countries’ ability to turn rhetoric into action by tracking learning.

Choose learning metrics based on what the country needs

When choosing which measures of learning to invest in, policy makers must consider the context. If assessment systems are nascent, priority should be given to fostering classroom assessment. Once that piece is in place, countries can develop relatively quick, sample-based, low-cost national assessments. When classroom and national assessments are established, much can be gained from participating in regional or global assessments that enable performance benchmarking. The ultimate goal is to build assessment systems in which different parts are aligned but serve different needs.

Not every student needs to be tested in national assessments. Sample-based assessments can accurately measure a system's performance. These assessments still require capable administrators, but they are much less expensive than census-based assessments. They can also be administered more often. Schools participating in these assessments do not have to be identified. This helps lower the stakes, making the assessments less susceptible to perverse responses by teachers or schools.

Assessment systems should test students at an age when effective remedial action remains possible. Of 121 countries in four regions, a third lack any reporting data on the reading and mathematics proficiency levels of children at the end of primary school.³⁹ Only

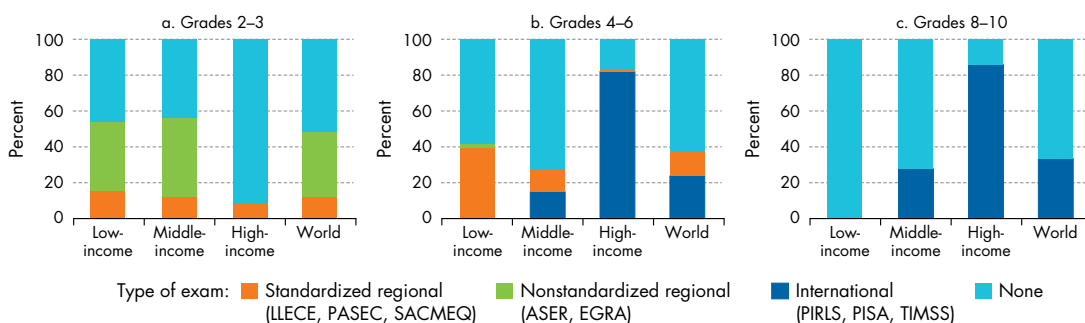
half of the countries surveyed produce data or participate in any regional or international tests to assess mathematics at the end of lower secondary school. Just under half assess reading. This means that comparable information about learning is missing for most children and youth outside of high-income countries (figure 4.1).⁴⁰

Will learning metrics narrow the vision for education?

Putting emphasis on measurable learning does not mean ignoring other outcomes of education, such as physical, moral, civic, or artistic development. Indeed, focusing on learning—and on the educational quality that drives it—is more likely to crowd in these other desirable outcomes. Conditions that allow children to spend two or three years in school without learning to read a single word or to reach the end of primary school without learning two-digit subtraction are not conducive to reaching the higher goals of education. An experiment in Andhra Pradesh, India, that rewarded teachers for gains in measured learning in math and language led to improved outcomes not just in those subjects, but also in science and social studies—even though there were no rewards for improvement in the latter two subjects.⁴¹ A study of ninth graders in the United States found that behavioral factors correlate positively with test scores.⁴² Another U.S. study revealed that teachers who improve test scores also improve broad outcomes into adulthood.⁴³

Figure 4.1 No internationally comparable data on learning are available for most children outside of high-income countries

Percentage of children in countries that have reported mathematics and reading scores since 2000 for ASER, EGRA, LLECE, PASEC, PIRLS, PISA, SACMEQ, and TIMSS, by income group



Source: WDR 2018 team, using data from Sandefur (2017). Data at http://bit.do/WDR2018-Fig_4-1.

Note: ASER = Annual Status of Education Report; EGRA = Early Grade Reading Assessment; LLECE = Latin American Laboratory for Assessment of the Quality of Education; PASEC = Programme d'Analyse des Systèmes Éducatifs de la Confemem; PIRLS = Progress in International Reading Literacy Study; PISA = Programme for International Student Assessment; SACMEQ = Southern and Eastern Africa Consortium for Monitoring Educational Quality; TIMSS = Trends in International Mathematics and Science Study.

Learning assessments of key foundational subjects such as language and mathematics are likely to be good proxies for whether an education system is delivering on its broad promise.

That said, cognitive skills are not the only skills that matter. Socioemotional skills (sometimes called noncognitive skills) such as grit, self-control, self-management, effective communication, and pro-social behavior can be central to not just economic outcomes but life outcomes more broadly.⁴⁴ Evidence from high-income countries suggests that such skills strongly affect employment status, work experience, occupational choice, and wages.⁴⁵ They also reduce risky behaviors such as crime, violence, or drug use.⁴⁶ For example, a study from the United Kingdom found that even after controlling for cognitive skills, socioemotional skills were important for predicting whether individuals stayed in school, obtained a degree, were employed, smoked, or were involved in a crime.⁴⁷ An understanding of how to measure these skills, along with how to influence them, is growing rapidly.⁴⁸ Like cognitive skills, socioemotional skills develop early in life but are malleable.⁴⁹ In fact, socioemotional skills help build cognitive skills and vice versa, with current skill levels dependent on investments made earlier in life (see spotlight 3).⁵⁰

Lower-performing countries probably do not face the same sharp trade-offs faced by high-performing countries on the education frontier. Economists use the concept of the production possibilities frontier

to understand how producers—or in this case, countries—make trade-offs between production of different goods (figure 4.2). For example, in recent years many stakeholders in the Republic of Korea have argued that their high-performing education system places too much emphasis on test scores (shown in figure 4.2 as “measured learning”) and not enough on creativity or certain socioemotional skills such as teamwork (“other outputs”). Implicitly, this Korean debate is about whether to try to move up and to the left on the frontier—that is, from A toward B. But in the low-learning trap, represented by “low-performing country C” in the figure, there is so much slack that this OECD-driven debate is not relevant. Country C has an opportunity to improve on both measured learning and other education outputs at the same time.

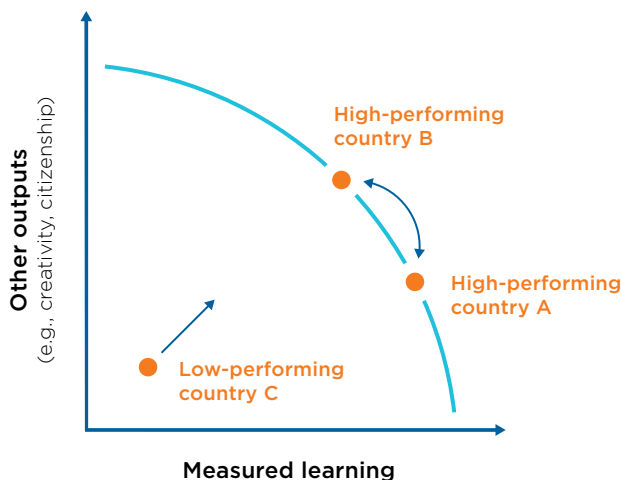
Six tips for effective learning measurement

Tip 1: Measure gaps. The learning crisis will be truly salient politically only when vulnerable subpopulations, who are disproportionately likely to suffer from learning gaps, are adequately covered by national assessment systems. To ensure that happens, assessments should be deployed in a way that shines a light on all children. Measurement must allow for the disaggregation of data around important dimensions such as socioeconomic status, gender, location, or disability status. In particular, groups at risk for social or economic exclusion may need to be oversampled to ensure adequate representation.⁵¹

Tip 2: Track progress. The use of uniform methodologies, approaches, and psychometrics across years is crucial for education systems to discern trends in learning over time and changes in learning gaps across tests. Year-on-year comparisons of learning progress should also be ensured for vulnerable subpopulations.

Tip 3: Test students when effective action is still possible. Returns from student assessments will be maximized if they focus on ensuring that students attain basic skills—literacy, numeracy, critical thinking—early in their schooling. Systems should also consider household-based testing, which would allow assessments to cover students not currently in school, making the resulting measurement more useful for universal learning targets. Household testing would also allow more nuanced understanding of all the different influences on a child’s school access and learning outcomes. To that end, standardized learning modules can be included at little additional cost

Figure 4.2 Low-performing countries don’t face sharp trade-offs between learning and other education outputs



Source: WDR 2018 team.

in surveys conducted both nationally (such as income and consumption surveys) and internationally (such as Living Standards Measurement Study surveys or Demographic and Health Surveys).

Tip 4: Balance the stakes. No single measure should be misused or overused. One way to avoid that outcome is to frame learning measures that guide policy as low-stakes diagnostic tools—not as one summary number that determines sanctions and rewards. Again, “learning metrics” should be considered a system of tools, each with its own place and purpose.⁵²

Tip 5: Good design is not enough—facilitate action. Learning measures should be used explicitly not just for tracking progress, but also for policy making.⁵³ One way to ensure that happens is to devote resources (including effort) to the timely distribution

of understandable results to key stakeholders. Another factor is an open, collaborative process for instrument design. Student assessments developed with the collaboration of various stakeholders are more likely to be considered valid and relevant at local levels.

Tip 6: Exploit global public goods on learning. Leveraging international assessments can yield high returns. For example, there is considerable advantage to forging common links between international and regional assessments so they can be put on the same scale. This not only increases harmonization between international assessments such as PISA and TIMSS, but also allows ties to national and citizen-led assessments, enabling meaningful global tracking (box 4.2). Researchers have tried to link various assessments after the fact, but these attempts have faced severe

Box 4.2 A global learning metric?

A global learning metric could help bring learning center stage, making it more salient. Such a metric would use an internationally comparable scale to consistently track progress and identify gaps across contexts. It would enable comparisons across children, households, schools, and locations.

Beyond its technical dividends, a global metric would motivate action and generate accountability for learning. By showing what is possible, it could point to what countries should be aspiring to—and create pressure to meet those aspirations. By benchmarking learning gaps among disadvantaged groups, a global metric could also create pressures for social mobility within countries. Furthermore, comparable learning data could increase the effectiveness of global research, international partnerships, and global aid for learning. Such data could also help countries develop their capacity for analyzing results to drive policy.

To be sure, there are technical and political challenges that would go hand in hand with adopting a global metric. The first is how to generate a global consensus on the metric’s scope. A global metric would require making choices about approach, target sample, and interpretation, which could prove controversial. In addition, challenges would arise with financing, implementation capacity, and political will. Many developing countries lack infrastructure for data collection, organization, analysis, and mechanisms to provide feedback to educators, parents, or communities. These are all necessary ingredients for turning metrics into action.

But most of these problems are surmountable. Global advocacy is generating sound technical recommendations on what a global metric could look like. Although there are no agreed-on standards of proficiency and no agreed-on tests to ensure that countries’ measures of learning are comparable to each other and over time, several global initiatives—such as the Global Alliance to Monitor Learning, the Assessment for Learning (A4L) initiative, and the International Commission on Financing Global Education Opportunity—are generating momentum. Other challenges could be overcome through clear goals and quality thresholds. A global metric can succeed only if it is explicitly framed as a complement to national assessment systems—not as a substitute for them. In fact, information from the global metric could be used to strengthen the capacity of national systems.

The political will needed for a global metric might be easier to mobilize if the needs of developing countries are prioritized and the metric’s advantages are clearly communicated. Estimates suggest that only 3 percent of official development assistance for education is spent on global public goods such as data and research; for health, that share is 20 percent.^a Returns from investing more on education data could be enormous if they help focus attention on ensuring that students attain basic skills in their early years.

Source: WDR 2018 team.

a. Schäferhoff and Burnett (2016).

technical challenges.⁵⁴ Ex ante linking of measurements through common items is likely to prove much more technically sound and cost-effective.

* * *

Education systems are unlikely to tackle the learning crisis unless it becomes clearly visible. This is possible only through well-designed measures of learning. To

be effective, “learning metrics” must overcome two important challenges: ensuring that information leads to action, and minimizing the potential perverse impacts of measurement. Alarm at the rise of a “testing” culture has dominated recent discourse. But in most low-learning contexts there is too little assessment and, consequently, too little accountability for learning in the system.

Notes

1. Pritchett, Banerji, and Kenny (2013).
2. Muralidharan, Singh, and Ganimian (2016).
3. Eggen and Stobart (2014); Sellar and Lingard (2013).
4. Carnoy and others (2016).
5. Pritchett (2013).
6. OECD (2011).
7. Guimarães de Castro (2012).
8. Breakspear (2012).
9. Kulpoor (1998).
10. Ertl (2006).
11. Greaney and Kellaghan (2008).
12. Dee and Jacob (2011).
13. Booher-Jennings (2005); Cullen and Reback (2006); Figlio and Getzler (2006); Jacob (2005); Jennings and Beveridge (2009); Neal and Schanzenbach (2010); Reback (2008).
14. Carnoy and others (2016); Glewwe and others (2017); OECD (2016); Xu and Dronkers (2016).
15. Székely (2011).
16. Greaney and Kellaghan (2008).
17. Baker and others (2010); Dixon and others (2013).
18. Bellamy and Raab (2005); Meijer (2009).
19. Bruns, Filmer, and Patrinos (2011).
20. Afrobarometer (2015); Uwezo (2014).
21. World Bank (2003).
22. Pritchett, Banerji, and Kenny (2013).
23. R4D (2015).
24. ASER Centre (2016).
25. Lieberman, Posner, and Tsai (2014).
26. Michener and Ritter (2016); Tanaka (2001).
27. Nicolai and others (2014).
28. Ganimian (2015).
29. Fox (2007); Hood (2010); Worthy (2015).
30. Meckes and Carrasco (2006).
31. Benveniste (2002); Peterson and West (2003).
32. Loewenstein, Sunstein, and Golman (2014).
33. Björkman and Svensson (2010).
34. Barr and others (2012); Björkman and Svensson (2010).
35. Barr and others (2012); Lieberman, Posner, and Tsai (2014).
36. Breakspear (2012).
37. Figazzolo (2009).
38. Tawil and others (2016).
39. UIS (2016).
40. By linking items across assessments—and including national assessments—coverage might be increased. But the fact that several large low- and middle-income

- countries still lack measures of learning means that comparable information remains missing for many children and youth outside of high-income countries.
41. Muralidharan and Sundararaman (2011).
42. Jackson (2016).
43. Chetty and others (2010).
44. Durlak and others (2011); Heckman, Pinto, and Savelyev (2013); Murnane and others (2001).
45. Heckman, Stixrud, and Urzua (2006).
46. Durlak, Weissberg, and Pachan (2010).
47. Carneiro, Crawford, and Goodman (2007).
48. Carneiro, Crawford, and Goodman (2007); Heckman, Pinto, and Savelyev (2013).
49. Heckman, Stixrud, and Urzua (2006).
50. Cunha and Heckman (2007, 2008); OECD (2015).
51. Sandefur (2016).
52. Neal (2013).
53. Guimarães de Castro (2012).
54. Altinok, Diebolt, and Demeulemeester (2014); Altinok and Murseli (2007); Sandefur (2017).

References

- Afrobarometer. 2015. “Uganda, Round 6 Data (2015).” Democracy in Africa Research Unit, Center for Social Science Research, University of Cape Town, Rondebosch, South Africa. <http://afrobarometer.org/data/uganda-round-6-data-2015>.
- Altinok, Nadir, Claude Diebolt, and Jean-Luc Demeulemeester. 2014. “A New International Database on Education Quality: 1965–2010.” *Applied Economics* 46 (11): 1212–47.
- Altinok, Nadir, and Hatidje Murseli. 2007. “International Database on Human Capital Quality.” *Economics Letters* 96 (2): 237–44.
- ASER Centre. 2016. “Annual Status of Education Report 2016.” New Delhi. <http://www.asercentre.org/p/289.html>.
- Baker, Eva L., Paul E. Barton, Linda Darling-Hammond, Edward Haertel, Helen F. Ladd, Robert L. Linn, Diane Ravitch, et al. 2010. “Problems with the Use of Student Test Scores to Evaluate Teachers.” EPI Briefing Paper #278, Economic Policy Institute, Washington, DC.
- Barr, Abigail, Frederick Mugisha, Pieter Serneels, and Andrew Zeitlin. 2012. “Information and Collective Action in Community-Based Monitoring of Schools: Field and

- Lab Experimental Evidence from Uganda.” Working paper, Georgetown University, Washington, DC.
- Bellamy, Christine, and Charles Raab. 2005. “Joined-Up Government and Privacy in the United Kingdom: Managing Tensions between Data Protection and Social Policy, Part II.” *Public Administration* 83 (2): 393–415.
- Benveniste, Luis. 2002. “The Political Structuration of Assessment: Negotiating State Power and Legitimacy.” *Comparative Education Review* 46 (1): 89–118.
- Birdsall, Nancy, Barbara Bruns, and Janeen Madan. 2016. “Learning Data for Better Policy: A Global Agenda.” CGD Policy Paper, Center for Global Development, Washington, DC. <http://www.cgdev.org/sites/default/files/learning-data-better-policy.pdf>.
- Björkman, Martina, and Jakob Svensson. 2010. “When Is Community-Based Monitoring Effective? Evidence from a Randomized Experiment in Primary Health in Uganda.” *Journal of the European Economic Association* 8 (2–3): 571–81.
- Booher-Jennings, Jennifer. 2005. “Below the Bubble: ‘Educational Triage’ and the Texas Accountability System.” *American Educational Research Journal* 42 (2): 231–68.
- Breakspear, Simon. 2012. “The Policy Impact of PISA: An Exploration of the Normative Effects of International Benchmarking in School System Performance.” OECD Education Working Paper 71, Organisation for Economic Co-operation and Development, Paris.
- Bruns, Barbara, Deon Filmer, and Harry Anthony Patrinos. 2011. *Making Schools Work: New Evidence on Accountability Reforms*. Human Development Perspectives Series. Washington, DC: World Bank.
- Carneiro, Pedro, Claire Crawford, and Alissa Goodman. 2007. “The Impact of Early Cognitive and Non-cognitive Skills on Later Outcomes.” CEE Discussion Paper 0092, Centre for the Economics of Education, London School of Economics.
- Carnoy, Martin, Tatiana Khavenson, Prashant Loyalka, William H. Schmidt, and Andrey Zakharov. 2016. “Revisiting the Relationship between International Assessment Outcomes and Educational Production: Evidence from a Longitudinal PISA-TIMSS Sample.” *American Educational Research Journal* 53 (4): 1054–85.
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. 2010. “How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star.” NBER Working Paper 16381, National Bureau of Economic Research, Cambridge, MA.
- Cullen, Julie Berry, and Randall Reback. 2006. “Tinkering toward Accolades: School Gaming under a Performance Accountability System.” In *Improving School Accountability*, edited by Timothy J. Gronberg and Dennis W. Jansen, 1–34. Advances in Applied Microeconomics Series 14. Bingley, U.K.: Emerald Publishing Limited.
- Cunha, Flavio, and James J. Heckman. 2007. “The Technology of Skill Formation.” IZA Discussion Paper 2550, Institute for the Study of Labor, Bonn, Germany. <http://nbn-resolving.de/urn:nbn:de:10:1-20080425464>.
- . 2008. “Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation.” *Journal of Human Resources* 43 (4): 738–82.
- Dee, Thomas S., and Brian Jacob. 2011. “The Impact of No Child Left Behind on Student Achievement.” *Journal of Policy Analysis and Management* 30 (3): 418–46.
- Dixon, Ruth, Christiane Arndt, Manuel Mullers, Jarmo Vakkuri, Kristiina Engblom-Pelkkala, and Christopher Hood. 2013. “A Lever for Improvement or a Magnet for Blame? Press and Political Responses to International Educational Rankings in Four EU Countries.” *Public Administration* 91 (2): 484–505.
- Durlak, Joseph A., Roger P. Weissberg, Allison B. Dymnicki, Rebecca D. Taylor, and Kriston B. Schellinger. 2011. “The Impact of Enhancing Students’ Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions.” *Child Development* 82 (1): 405–32.
- Durlak, Joseph A., Roger P. Weissberg, and Molly Pachan. 2010. “A Meta-Analysis of After-School Programs That Seek to Promote Personal and Social Skills in Children and Adolescents.” *American Journal of Community Psychology* 45 (3–4): 294–309.
- Eggen, Theo J. H. M., and Gordon Stobart, eds. 2014. *High-Stakes Testing in Education: Value, Fairness, and Consequences*. New York: Routledge.
- Elwick, Alex, and Tony McAleavy. 2015. *Interesting Cities: Five Approaches to Urban School Reform*. Reading, U.K.: CfBT Education Trust.
- Ertl, Hubert. 2006. “Educational Standards and the Changing Discourse on Education: The Reception and Consequences of the PISA Study in Germany.” *Oxford Review of Education* 32 (5): 619–34.
- Figazzolo, Laura. 2009. “Impact of PISA 2006 on the Education Policy Debate.” Working paper, Education International, Brussels.
- Figlio, David N., and Lawrence S. Getzler. 2006. “Accountability, Ability, and Disability: Gaming the System?” In *Improving School Accountability*, edited by Timothy J. Gronberg and Dennis W. Jansen, 35–49. Advances in Applied Microeconomics Series 14. Bingley, U.K.: Emerald Publishing Limited.
- Fox, Jonathan. 2007. “The Uncertain Relationship between Transparency and Accountability.” *Development in Practice* 17 (4–5): 663–71.
- Ganimian, Alejandro J. 2015. *El Termómetro Educativo: Informe Sobre el Desempeño de Argentina en los Operativos Nacionales de Evaluación (One) 2005–2013*. Buenos Aires: Proyecto Educar 2050.
- Glewwe, Paul W., Jongwook Lee, Khoa Vu, and Hai-Anh H. Dang. 2017. “What Explains Vietnam’s Exceptional Performance in Education Relative to Other Countries? Analysis of the 2012 PISA Data.” Paper presented at the RISE Annual Conference, Center for Global Development, Washington, DC, June 15–16.
- Greaney, Vincent, and Thomas Kellaghan. 2008. *Assessing National Achievement Levels in Education*. Vol. 1, *National Assessments of Educational Achievement*. Washington, DC: World Bank.
- Guimarães de Castro, Maria Helena. 2012. “Developing the Enabling Context for Student Assessment in Brazil.” SABER Student Assessment Working Paper 7, Systems Approach for Better Education Results, World Bank, Washington, DC.

- Heckman, James J., Rodrigo Pinto, and Peter Savelyev. 2013. "Understanding the Mechanisms through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review* 103 (6): 2052–86.
- Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3): 411–82.
- Holbein, John. 2016. "Left Behind? Citizen Responsiveness to Government Performance Information." *American Political Science Review* 110 (2): 353–68.
- Hood, Christopher. 2010. *The Blame Game: Spin, Bureaucracy, and Self-Preservation in Government*. Princeton, NJ: Princeton University Press.
- Jackson, C. Kirabo. 2016. "What Do Test Scores Miss? The Importance of Teacher Effects on Non-test Score Outcomes." NBER Working Paper 22226, National Bureau of Economic Research, Cambridge, MA.
- Jacob, Brian A. 2005. "Accountability, Incentives, and Behavior: The Impact of High-Stakes Testing in the Chicago Public Schools." *Journal of Public Economics* 89 (5): 761–96.
- Jennings, Jennifer L., and Andrew A. Beveridge. 2009. "How Does Test Exemption Affect Schools' and Students' Academic Performance?" *Educational Evaluation and Policy Analysis* 31 (2): 153–75.
- Kulpoo, Dhurumbeer. 1998. "The Quality of Education: Some Policy Suggestions Based on a Survey of Schools: Mauritius." SACMEQ Policy Research Report No. 1, United Nations Educational, Scientific, and Cultural Organization, Port Louis, Mauritius.
- Lieberman, Evan S., Daniel N. Posner, and Lily L. Tsai. 2014. "Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya." *World Development* 60: 69–83.
- Loewenstein, George, Cass R. Sunstein, and Russell Golman. 2014. "Disclosure: Psychology Changes Everything." *Annual Review of Economics* 6 (1): 391–419.
- Meckes, Lorena, and Rafael Carrasco. 2006. "SIMCE: Lessons from the Chilean Experience in National Assessment Systems of Learning Outcomes." Paper presented at World Bank and Inter-American Development Bank's Conference, "Lessons from Best Practices in Promoting Education for All: Latin America and the Caribbean," Cartagena de Indias, Colombia, October 9–11.
- Meijer, Albert. 2009. "Understanding Modern Transparency." *International Review of Administrative Sciences* 75 (2): 255–69.
- Michener, Gregory, and Otavio Ritter. 2016. "Comparing Resistance to Open Data Performance Measurement: Public Education in Brazil and the UK." *Public Administration* 95 (1): 4–21.
- Muralidharan, Karthik, Abhijeet Singh, and Alejandro J. Ganimian. 2016. "Disrupting Education? Experimental Evidence on Technology-Aided Instruction in India." NBER Working Paper 22923, National Bureau of Economic Research, Cambridge, MA.
- Muralidharan, Karthik, and Venkatesh Sundararaman. 2011. "Teacher Performance Pay: Experimental Evidence from India." *Journal of Political Economy* 119 (1): 39–77.
- Murnane, Richard J., John B. Willett, M. Jay Braatz, and Yves Duhaldeborde. 2001. "Do Different Dimensions of Male High School Students' Skills Predict Labor Market Success a Decade Later? Evidence from the NLSY." *Economics of Education Review* 20 (4): 311–20.
- Neal, Derek. 2013. "The Consequences of Using One Assessment System to Pursue Two Objectives." *Journal of Economic Education* 44 (4): 339–52.
- Neal, Derek, and Diane Whitmore Schanzenbach. 2010. "Left Behind by Design: Proficiency Counts and Test-Based Accountability." *Review of Economics and Statistics* 92 (2): 263–83.
- Nicolai, Susan, Leni Wild, Joseph Wales, Sébastien Hine, and Jakob Engel. 2014. "Unbalanced Progress: What Political Dynamics Mean for Education Access and Quality." ODI Development Progress Working Paper 5, Overseas Development Institute, London.
- OECD (Organisation for Economic Co-operation and Development). 2011. *Strong Performers and Successful Reformers in Education: Lessons from PISA for the United States*. Paris: OECD.
- . 2015. *Skills for Social Progress: The Power of Social and Emotional Skills*. Paris: OECD.
- . 2016. *PISA 2015 Results: Excellence and Equity in Education*. Vol. 1. Paris: OECD.
- Peterson, Paul E., and Martin R. West. 2003. *No Child Left Behind? The Politics and Practice of School Accountability*. Washington, DC: Brookings Institution Press.
- Pritchett, Lant. 2013. *The Rebirth of Education: Schooling Ain't Learning*. Washington, DC: Center for Global Development; Baltimore: Brookings Institution Press.
- Pritchett, Lant, Rukmini Banerji, and Charles Kenny. 2013. "Schooling Is Not Education! Using Assessment to Change the Politics of Non-learning." CGD Report, Center for Global Development, Washington, DC.
- R4D (Results for Development Institute). 2015. "Bringing Learning to Light: The Role of Citizen-Led Assessments in Shifting the Education Agenda." R4D, Washington, DC.
- Reback, Randall. 2008. "Teaching to the Rating: School Accountability and the Distribution of Student Achievement." *Journal of Public Economics* 92 (5): 1394–1415.
- Sandefur, Justin. 2016. "Internationally Comparable Mathematics Scores for Fourteen African Countries." CGD Working Paper 444 (December), Center for Global Development, Washington, DC.
- . 2017. "The Case for Global Standardized Testing." *Views from the Center: Education, Education Reform* (blog), April 27. <https://www.cgdev.org/blog/case-global-standardized-testing>.
- Schäferhoff, Marco, and Nicholas Burnett. 2016. "Rethinking the Financing and Architecture of Global Education." Background Paper: The Learning Generation (April 29), International Commission on Financing Global Education Opportunity, New York.
- Sellar, Sam, and Bob Lingard. 2013. "The OECD and Global Governance in Education." *Journal of Education Policy* 28 (5): 710–25.
- Székel, Miguel. 2011. "Toward Results-Based Social Policy Design and Implementation." CGD Working Paper 249, Center for Global Development, Washington, DC.

- Tanaka, Shinichiro. 2001. "Corruption in Education Sector Development: A Suggestion for Anticipatory Strategy." *International Journal of Educational Management* 15 (4): 158–66.
- Tawil, Sobhi, Margarete Sachs-Israel, Huong Le Thu, and Matthias Eck. 2016. *Unpacking Sustainable Development Goal 4 Education 2030—Guide*. Paris: United Nations Educational, Scientific, and Cultural Organization.
- UIS (UNESCO Institute for Statistics). 2016. "Laying the Foundation to Measure Sustainable Development Goal 4." Sustainable Development Data Digest, UIS, Montreal.
- Uwezo. 2014. "Are Our Children Learning? Literacy and Numeracy across East Africa 2013." Twaweza, Nairobi.
- World Bank. 2003. *World Development Report 2004: Making Services Work for Poor People*. Washington, DC: World Bank; New York: Oxford University Press.
- Worthy, Ben. 2015. "The Impact of Open Data in the UK: Complex, Unpredictable, and Political." *Public Administration* 93 (3): 788–805.
- Xu, Duoduo, and Jaap Dronkers. 2016. "Migrant Children in Shanghai: A Research Note on the PISA-Shanghai Controversy." *Chinese Sociological Review* 48 (3): 271–95.