

The Possibility and Desirability of Global Learning Metrics: Commentaries

The views expressed herein are those of the authors and do not necessarily reflect the views of CASGE.

The Possibility and Desirability of Global Learning Metrics: Commentaries

<http://dx.doi.org/10.14507/casge1.2017>

CASGE working papers are circulated for discussion and comment purposes. They have not been peer-reviewed.

copyright notice



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 3.0 Unported License. Readers are free to copy, display, and distribute content that appear in *CASGE* as long as the work is attributed to the author(s) and *CASGE*, it is distributed for non-commercial purposes only, and no alteration or transformation is made in the work. All other uses must be approved by the author(s) or *CASGE*. By submitting a manuscript, authors agree to transfer without charge the following rights to *CASGE* upon acceptance of the manuscript: first worldwide serial publication rights and the right for *CASGE* to grant permissions as its editors judge appropriate for the redistribution of the content, its abstract, and metadata associated with the working paper in professional indexing and reference services. Any revenues from such redistribution are used solely to support the continued publication and distribution of working papers.

Executive Summary

The Inaugural Symposium of the Comparative and International Education Society (CIES) was held from November 10-11, 2016 in Scottsdale, Arizona. The symposium was hosted by Arizona State University - Mary Lou Fulton Teachers College's Center for Advanced Studies in Global Education (CASGE) and edXchange. The event was titled "The Possibility and Desirability of Global Learning Metrics: Comparative Perspectives on Education Research, Policy and Practice." It brought together academics, practitioners, policymakers, educators, social activists, and others for an alternating series of keynote plenary debates and parallel sessions about the desirability and feasibility of global learning metrics. This working paper contains the position statements of many of the keynote plenary debate speakers, including David Edwards (Education International), Eric A. Hanushek (Stanford University), Monisha Bajaj (University of San Francisco), Aaron Benavot (UNESCO Global Monitoring Report and University at Albany-SUNY), Radhika Gorur (Deakin University), Pasi Sahlberg and Jonathan Hasak (University of Helsinki and Year Up), J. Douglas Willms (University of New Brunswick and The Learning Bar Inc), Supriya Baily (George Mason University) and Stafford Hood (University of Illinois at Urbana-Champaign).

Table of Contents

Introduction	5
Are Global Learning Metrics Desirable?	7
David Edwards, Education International	7
Eric A. Hanushek, Stanford University	10
Are Global Learning Metrics Feasible?.....	12
Monisha Bajaj, University of San Francisco.....	12
Aaron Benavot, UNESCO Global Monitoring Report, University at Albany-SUNY	15
Can Global Learning Metrics be Pedagogically Innovative?	17
Radhika Gorur, Deakin University	17
Pasi Sahlberg and Jonathan Hasak, University of Helsinki and Year Up	21
J. Douglas Willms, University of New Brunswick and The Learning Bar Inc.	24
Can Global Learning Metrics be Culturally Responsive?.....	26
Supriya Baily, George Mason University	26
Stafford Hood, University of Illinois at Urbana-Champaign	28
Acknowledgements	31

Introduction

Iveta Silova, Gustavo E. Fischman, and Janna Goebel, Arizona State University

The Inaugural Symposium of the Comparative and International Education Society (CIES) took place on November 10-11, 2016 in Scottsdale, Arizona. The Symposium was hosted by Arizona State University - Mary Lou Fulton Teachers College's Center for Advanced Studies in Global Education ([CASGE](#)) and [edXchange](#), with the generous support of the Open Society Foundations, the Comparative and International Education Society, and ASU International Development. The event brought together 151 individuals from 61 institutions, 17 different countries, and 17 states within the United States for an alternating series of keynote plenary debates and parallel sessions about the desirability and feasibility of global learning metrics.

We selected global learning metrics as the focus of the Symposium because it is a timely and increasingly challenging educational and political issue at the center of multiple global debates about the future of education. Learning outcomes have recently been enshrined as central policy objectives in the new international education and development agenda. Unlike goals that seek to universalize access for education, for which consensus is strong, debates around learning are considerably more contested. Proponents argue that more robust global learning metrics have the potential to reduce academic disparities and improve learning outcomes for children across different contexts. Critics note that such universal measures typically focus on a narrow assessment of basic skills, while overlooking the importance of a more holistic approach to education, including human rights, aesthetics, morality, religion, or spirituality. Others call attention to the dangers associated with the emergence of the data-fixated punitive accountability regimes, privatization and marketization of public education, and a growing disconnect between systems, actors, and larger pedagogic changes. Some critics warn that global learning metrics can contribute to enacting hegemonic neocolonial globalization. More broadly, the debate about the global learning metrics reveals an underlying tension in our field - a tension between the desire to replicate and scale up "best practices" (and an assumption that there is a global consensus on what constitutes "good" education), on the one hand, and the awareness about the importance of context, and deeply culturally contextualized education practice, on the other hand. Bringing a comparative perspective to the disjuncture between replicability and contextuality is one way our field can contribute to education research and practice broadly.

This raises the central questions, which guided the organization of this Symposium: Are global learning metric desirable and are they feasible? How can learning among children be measured and compared across diverse contexts and systems? Which learning domains should be assessed and why? How is learning revised or reframed for those who have less power or less "value" in the society in which they reside? How, if at all, are learning assessments actually used by governments, nongovernmental entities, teachers, curriculum developers, and other stakeholders? The Symposium brought together a group of researchers, policymakers, practitioners, and activists for a focused intellectual and policy engagement around these questions. While not designed to forge consensus or alignment, the Symposium was a step towards linking together academic research and policy debates in order to enable critical reflection, innovation, and proactive action in the area of developing global learning metrics.

The Symposium featured four moderated keynote debates, which addressed issues ranging from the desirability and feasibility of global learning metrics to their potential to be pedagogically innovative

and culturally responsive. The first plenary debate (moderated by Iveta Silova) focused on the different actors and rationales behind the development of global learning metrics, featuring insightful but heated exchanges of opinions between Silvia Montoya of the UNESCO Institute of Statistics, Karen Mundy from the Global Partnership for Education, Eric Hanushek from Stanford University, and David Edwards of Education International.

The second debate (moderated by Gustavo E. Fischman) addressed the issues of feasibility, especially in terms of measuring and comparing educational achievement across diverse contexts and educational systems. The panelists also discussed the possibility of balancing between the assessment of basic numeracy and literacy skills and the measurement of learning related to informational technologies, citizenship, human rights, sustainability, aesthetics, morality, religion and/or spirituality. The panelists included Monisha Bajaj from the University of San Francisco, Aaron Benavot from the UNESCO Global Monitoring Report, and David C. Berliner from ASU.

The third debate focused on the potential of global learning metrics to be pedagogically innovative. Moderated by Sherman Dorn from ASU, the panel brought together Chris Higgins from University of Illinois at Urbana-Champaign, Radhika Gorur of Deakin University, and Pasi Sahlberg from University of Helsinki to discuss the limitations of the big data movement and weigh its possible contributions to pedagogical innovation.

Finally, the fourth plenary debate (moderated by Gustavo E. Fischman) questioned the assumption that there is an agreement about what constitutes “good” and “quality” education worldwide. The plenary panel participants Supriya Baily from George Mason University, Stafford Hood from University of Illinois at Urbana-Champaign, Hugh McLean of the Open Society Foundations, and J. Douglas Willms of University of New Brunswick debated whether there is a global core of fundamental knowledge, skills and competencies that are relevant across different countries and discussed whether and how global learning metrics can capture the dynamics of race, ethnicity, class, gender, religion, and other factors that contribute to students’ cultural identities.

Are Global Learning Metrics Desirable?

**Are global learning metrics desirable?
That depends on what decision they are attempting to inform**

David Edwards, Education International

Over the last two decades there has been a steady drumbeat driving international education towards something commonly referred to as an outcomes-based approach. Its path of travel can be traced from the access or compensatory approach that focused on equalizing inputs through a quick blip in the land of the participatory approaches and process indicators and now locates itself squarely on the other side of the little black box, easily measured outcomes, targets and indicators.

We've finally arrived, some would say and rejoice that we can now hold ourselves accountable for results. Education is finally on par with its rich cousin health in terms of simple causal relations, but not attention, resourcing and evidence. Just do not ask us to question what is meant by "us" or "results."

Also, please do not ask for whose purpose and under whose terms, as that's clearly outside the bounds of polite post-access and process decorum. If you dare you will be immediately labeled a proponent of seat time and the status quo. In fact, in the negotiation of the latest round of Education For All (EFA) - round 3 where we merged and extended our agenda with the parallel Millennium Development Goals (MDGs) and Sustainable Development Goals (SDGs) (while simultaneously moving the goal posts to 2030), the outcome advocacy community (don't ask me to quantify the community) was clear it only wanted easy-to-measure, comparable targets with a simple to communicate goal.

The simplemetrification of EFA (Fischman, 2015) was on full display in the years prior to forming the Post 2015 High Level Panel. One early memory where the split was tangible came during the Global Campaign for Education's (GCE) World Assembly in February of 2011 when there was a push to get the right-based civil society organizations to get behind a single, early grade reading metric as their *raison d'être*. GCE's main donor even pulled its support when they had the audacity to continue pushing a broader rights approach with a notion of quality that included inputs, process and relevant outcomes in context. The multidimensional aspects of the word quality were particularly hard for the simplemetrification community who much preferred to appropriate the word learning (what others might call testing) and call everything else irrelevant.

These were the early days of global learning metrics within the broader EFA agenda. Of course, Progress in International Reading Literacy Study, Trends in International Mathematics and Science Study and Programme for International Student Assessment had been around a lot longer. Some were governed by ministers and others by agencies. Regional ones like the Laboratorio Latinoamericano de Evaluación de la Calidad de la Educación or Proyecto Regional de Indicadores Educativos in Latin America or Programme d'analyse des systèmes éducatifs de la confemen and Southern and Eastern Consortium for Monitoring Educational Quality in Africa were reporting bi or triannually to regional ministerial processes. Rankings were at the start of their global rage to 'simplymetrify' complex systems for busy bureaucrats and policymakers. Occasionally, a country like Cuba would dare to top the charts and United Nations Educational, Scientific and Cultural Organisation-Oficina Regional de Educación para América Latina y el Caribe would be asked to run their numbers again because after all, it was Cuba.

These were not seen by most teachers, parents and students as the best of times. In the US, data dashboards, high-stakes test accountability and data driven evangelists demanded 'no excuses' accountability for outcomes and a bit less so for school choice advocates.

Private companies that once published textbooks were now expanding their market share by developing the tests and test prep software at all levels. This small group of publishers was recently referred to as "the cartel" by the Organisation for Economic Co-operation and Development in their recent report on the Educating for Innovation and Innovating for Education (2016), because they controlled the lion's share of the testing and publishing market. A few of the biggest ones were able to win contracts to implement and then roll-up recommendations across national and global tests. They joined forces with (or spun off?) global consulting firms to produce equally simplified policy advice of things like "teacher selectivity matters" or "teach phonics" or "have an unrelenting focus on outcomes." The same firms would also alert investors when elections favored pro-privatization interest.

This is unsettling to people working in education over the past few decades. The reading comprehension challenge was a priority well before the learning crisis and decades of research and work was no less relevant because it wasn't funded by Department for International Development and United States Agency for International Development.

The focus on simple effect size over interactions and multiple measures coupled with the loud demand from education funders that we learn more from health and find our bed net, vaccine for illiteracy and easily communicable indicator meant the global snake oil peddlers, or entrepreneurs, returned to this now potentially profitable sector in droves.

Enter a group of self-appointed experts and think tanks, appeared to partner with testing companies and the education development industry to break reading down to its component parts, ignore broader social approaches and other such distractions so that it was an individual endeavor where a person followed a prescriptive script, tested, decoded and timed the whole thing with a stopwatch.

In response to the education community's refusal to implement their wishes, the simplemetrified donor community went looking for local groups in developing countries who could collect data to expose the level of school failure in league tables and reports. While not all citizen-led assessments took money from the same group, many did and the end result was predictable. These groups delighted testing companies, the IFIs and donor agencies who could finally force governments to meet their conditions from a demand standpoint. Instead of inputs and infrastructure and trained teachers, now grants and loans could be conditioned on countries taking part in global and regional evaluations, which were rather expensive and meant less money for schools and teachers.

In an important article written by Barrett and Sorensen (2015) the authors noted

A clear lesson is that indicators that are only partially fit-for-purpose in terms of how well they capture the meaning of the target can come to displace the parent target. Statistically robust indicators of what is readily measurable are very often only partially fit-for-purpose. Some measurement experts have suggested using them in combination with less robust indicators but more fit-for-purpose targets to construct a set of indicators that together are fit-for-purpose. Being fit-for-purpose in practice, however, will also depend on monitoring and

reporting mechanisms that amplify qualitative indicators, which tend to attract less attention. Indicators are more likely to have traction if they are comprehensible to and valued by educational professionals and civil society advocates of EFA, and can harness the support of wider society. This is called “communicability” or “salience.” (Langford 2012: 20)

Yet salience for educational professionals would have to wait because in a simplemetrified world of effect sizes and continuously recomputed rates of return, countries could be encouraged to confuse causality for correlation. Expert advice from economists made headlines. Poorly trained teachers get as bad results as untrained, poorly paid volunteers so you might as well go for cheaper teachers and spend the savings on ICT, multiyear testing and expensive advice from the usual suspects? (Only that’s exactly the opposite advice being given to OECD governments.) Politicians also saw value in so far that the political economy of education workers often meant it was difficult to enact sweeping changes that could limit the political power of organized labor and particularly influential professionals who could be found in almost every community.

For those of us who work with teachers, students and communities at a time of obscene inequality, vulnerability and violence one may forgive us for being less than enthusiastic about big data or a data revolution. We don't see summative, global learning metrics as a way out of poverty or an enabler of mass wisdom and social responsibility for our planet. Yet, we are in favor of metrics that fit our collectively agreed targets. We support the need for robust and fulsome evaluations of impact and particularly those that reveal inequality.

But, we work off the contention that one measures something or evaluates something to inform a decision. In the world of "small data" teachers make myriad instructional and interpersonal decisions daily. They do this on the basis of formative assessments and when they have time and are not administering tests and sorting students for a big data obsessed machine, they may even collaborate, share and plan with colleagues. That is of course if incentives have not been put in place to sanction them individually on the basis of a student’s test score.

There are a number of theories of change competing for relevance and prominence in this debate about global learning metrics. Some believe countries improve when they're embarrassed and ranked. Some believe that a simple metric reduces the ability to absorb complexity and maintain integrity, both features of high performing systems. Others wonder if the entire push for simplemetrification is not just another attempt by a market looking for an economy of scale and a simplified model of delivery...or deliverology.

So what is the decision at the global level we hope to answer with a global learning metric? Is it to address the issue of 123 million children outside of school and the ever-widening financing gap that helps keep them and the next generation out too? Is it to redistribute wealth and opportunities within a system that has already been designed to advantage certain countries and populations? Is it to help donors show impact to skeptical politicians and finance ministers with powerful stats? Or is it to develop a parallel system of expectations and options between rich and poor nations?

Something tells me it may not be just about getting real time information into the hands of well supported and trained education professionals working in quality environments and giving them the time and resources to make a difference in the lives of their students.

Are global learning metrics desirable?

Eric A. Hanushek, Stanford University

We have ample evidence about the power of international goals to focus attention on desirable outcomes and to promote improvement. The Millennium Development Goals (MDGs), calling for universal completion of lower secondary schooling, saw dramatic improvements in schooling access around the world. Unfortunately, many induced to attend did not learn anything – effectively erasing the potential gains coming from more education.

The recently ratified Sustainable Development Goals (SDGs) call for improvements in quality education, but without clearly defining “quality,” we could end up with similar disappointments to those accompanying the MDGs. The disappointment comes from the fact that the economic benefits of human capital development are directly related to the skills that are acquired.

Ludger Woessmann and I have investigated the impact of skills as measured by international tests on the earnings of individuals and on the growth of nations. In terms of economic growth, we find that performance on PISA and TIMSS tests is a strong predictor of economic growth (Hanushek and Woessmann (2015a)). Moreover, the analysis indicates that this can plausibly be thought of as a causal relationship.

The relationship is very strong, indicating that improvement in achievement (which in the aggregate we call the knowledge capital of nations) has large impacts on economic growth. This historic relationship also permits us to estimate how the MDGs or the SDGs could change the economic well-being of nations. Figure 1 summarizes our estimates of how reaching “basic skill” level – which we define as Level 1 on the PISA tests – would affect growth of GDP for the 76 countries with data. (Hanushek and Woessmann (2015b)). We summarize the present value of added growth over the next 80 years (the life span of somebody born today). We show the impacts for aggregations of countries grouped by the level of today’s GDP and consider: (1) simple quality improvement with no increased access to lower secondary schooling; (2) movement to full access to lower secondary schooling at each country’s current quality; and (3) universal access with students reaching the basic skill level.

The results are quite startling. For the lowest income countries that we observe, bringing current students up to basic skills would on average yield a present value of the growth dividend that was over six times current GDP. Just providing full access to the current schools would yield only twice current GDP, while universal basic skills would by historical standards produce a growth dividend of twelve times current GDP. Countries not currently tested would likely obtain larger gains. The goal of universal basic skills can be applied to all countries. While the most developed countries are close to complete access to lower secondary schooling, some students are left behind. In the U.S. this percentage is 23 percent, and bringing them up to Level 1 on PISA would imply over 1.5 times current US GDP. High income non-OECD includes many Arab oil countries, gaining even more.

There are also large differences around the world in the individual returns to skills as measured by standardized math tests. Recent estimates show that the returns to skills in different countries are closely related to the amount of change in the country as indexed by the growth rate in GDP (Hanushek et al., 2016). Said differently, people without skills are severely harmed if the country grows fast.

My proposed learning metrics would relate directly to the math, science, and reading skills that have been successfully measured in PISA and TIMSS. Of course, the application of this idea is not entirely simple, because the lowest income countries currently find that these test items are too difficult for their students. Thus, as a practical approach I would recommend developing tests that were informative of student differences (such as found in the regional tests of TERCE and SACMEQ) but also including linking items to PISA and TIMSS. Making learning a real goal would have large economic effects around the world.

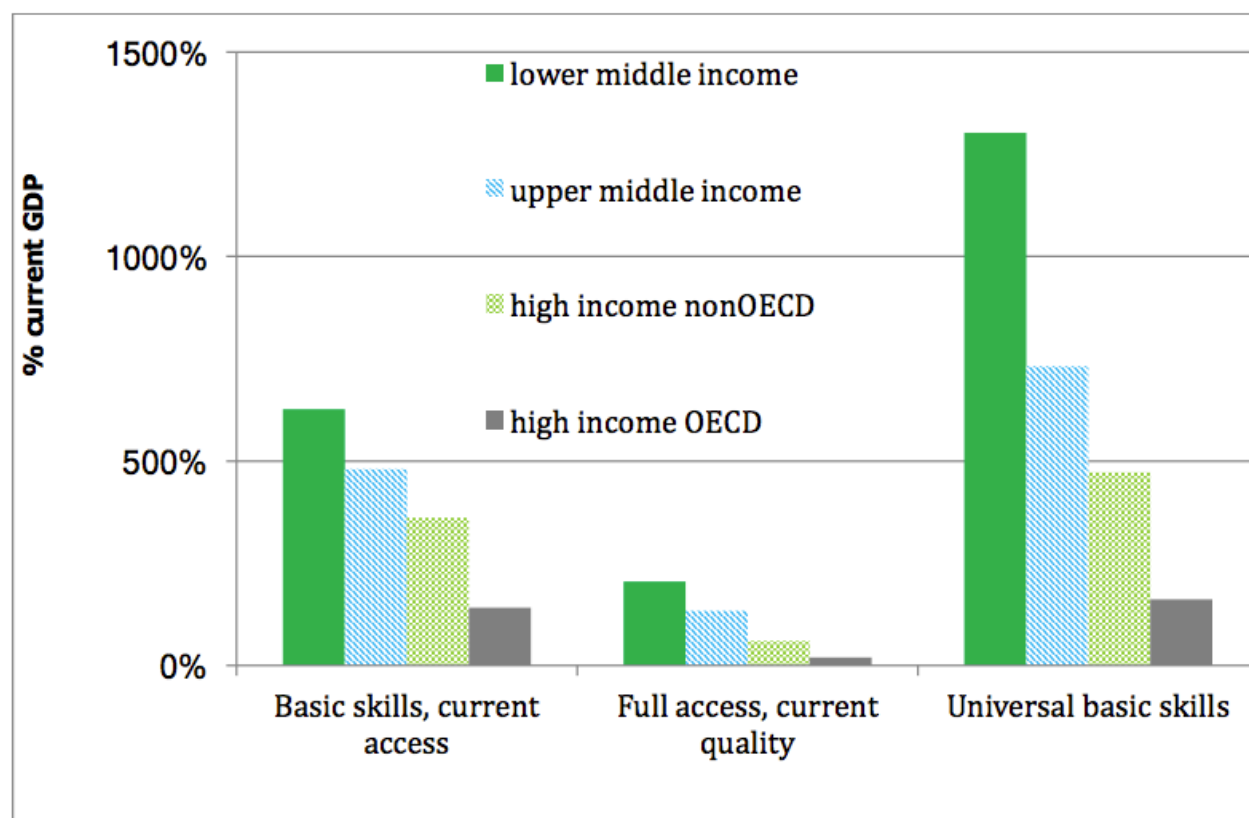


Figure 1. Increased GDP from Improved Skills, Improved Access, and Universally Higher Skills by Country Income Level (Source: Hanushek and Woessmann (2015b))

References

- Hanushek, Eric A., Guido Schwerdt, Simon Wiederhold, and Ludger Woessmann. 2016. "Coping with change: International differences in the returns to skills." NBER Working Paper No. 22657. Cambridge, MA: National Bureau of Economic Research (September).
- Hanushek, Eric A., and Ludger Woessmann. 2015a. *The knowledge capital of nations: Education and the economics of growth*. Cambridge, MA: MIT Press.
- Hanushek, Eric A., and Ludger Woessmann. 2015b. *Universal basic skills: What countries stand to gain*. Paris: Organisation for Economic Co-operation and Development.

Are Global Learning Metrics Feasible?

Are global learning metrics feasible? Beyond Measure: Valuing Human Rights in Global Education

Monisha Bajaj, University of San Francisco

How do we measure agency? How do we compare empathy? How do we evaluate respect for human rights? These are core components of a quality and comprehensive education to prepare young people for lifelong active citizenship, yet prove difficult to assess and measure in international learning metrics. Human rights education (HRE), in its ideal form, offers transformative learning that fosters agency, critical analysis, and social action. Many efforts towards including human rights information in textbooks or teacher education courses have been made across the globe; however, superficial forms of purely content-based reforms often miss the larger goal of HRE that include cognitive, affective and action-oriented dimensions of the holistic educational project (Bajaj, 2012; Bajaj et al., 2016; Flowers, 2003; Tibbitts, 2002).

Content. While the rise of human rights content in textbooks internationally signals the increasing popularity of rights discourses (Meyer et al., 2010), dynamic grassroots struggles for rights often get flattened and simplified when included in textbooks: Martin Luther King Jr.'s wide-ranging calls for economic and racial justice, anti-war stances, and radical nonviolence get reduced into four words—"I have a dream"—and a photo in textbooks whether in the United States, India or Taiwan. Thus, while including human rights content can be a starting point for students' continued exploration, on its own, the value of such content is limited if it fails to offer a connection between a historic struggle and students' own lives.

Attitudes. The 2016 cycle of the International Civic and Citizenship Study of the International Association for the Evaluation of Educational Achievement (or the IEA) has many indicators that offer 14-year olds across the world a chance to self-report on their attitudes towards human rights; this builds on previous inclusion of rights-related questions on previous IEA studies. Such inclusion is laudable and indeed offers valuable cross-national information about attitudes towards political rights, women's rights, and the rights of minority groups. And the mixed and varied performance on previous IEA studies on these indicators—related to, for example, attitudes towards immigrants and gender equality—suggests much more need for human rights education in school systems near and far.

Action. In my research in India and the United States, there are countless examples of human rights actions that would never be captured on an examination whether at one's school, a state-wide exam, or through a cross-national comparison. Yet, such experiences, in the cases of Fatima and Zau narrated below, often mark their learning experiences, offering critical learning moments that can prove transformative in the course of their lives.

I met Fatima in 2009 in her village in southern India. As the eldest child in her family, she was to be killed in the illegal but still common practice of female infanticide. Her grandmother intervened to save her, and in doing, was given the child to raise on her own with a meager salary as a sweeper at a local school. Fatima's attendance in school was sporadic with chores to do and her performance less than average when there. After a non-governmental organization (NGO), the Institute of Human Rights Education, introduced a three-year course in human rights education, Fatima had a chance to

explore poetry and public speaking through competitions and events held in collaboration with her school and the NGO. Although a high-stakes exam after the 10th grade in India tells students what careers they can and can't pursue based on their scores, even after the exam, Fatima was passionate about human rights, had written over 100 poems about them, and her teachers were helping her publish them to raise money for her college costs (Bajaj, 2012).

I met Zau in 2014 when he was a student in a human rights club my research team ran for newly arrived immigrant and refugee youth in California. Zau was a refugee from Burma, a fifth year senior who was simultaneously taking classes at community college and still at high school because he hadn't yet successfully passed the state-wide high school exit examination. Although Zau had just been in the U.S. a few years, the state only offered the exam in English and no accommodations were made for English Language Learners. When Zau was 10, his family's small plot and house were confiscated by the Burmese military, and he was forced to go to work to support his family until they could flee to Malaysia and later the U.S. In the human rights club, when discussing child labor, Zau shared,

That's like my story. I had to go find gold in the mines when I was 11 years old. We had to dig holes 20 feet down, and then go inside to see if there was gold. It was so quiet when I would go inside there. Because the government had taken our farms over and built things on our land, we had to leave and look for work. The mine owners hire children because we have more energy and we are small so we can go inside the holes and go way down into the mines. Plus, if something falls as we dig, we can move quickly to escape. It was really scary, but I didn't really have a choice but to do it. (as cited in Bajaj et al., 2017)

Considered a "failure" by state-wide standards, Zau found resonance in learning about human rights and other children forced to work; he said it made him feel better that "it's not just us" going through these hardships, and that there are people working towards ending such violations. This resonates with Varenne and McDermott's (1998) notion of "successful failure" in which school evaluations prove partial and insufficient in recognizing or valuing the strengths and potential of students who fall outside the metrics of rigidly-defined success.

The experiences of Fatima and Zau suggest that key components and forms of education towards individual self-realization and the social good—critical consciousness, agency, promoting sustainability, respecting human rights, co-existing peacefully, among others—defy assessment and evaluation: they are indeed, beyond measure.

Human rights education praxis requires more comprehensive and long-term modes of valuation. We cannot presume a priori the outcomes for the destination of human rights education given that each student will internalize and make sense of it differently based on identity, social location, and historic and current forms of privilege or marginalization. Further, decontextualized human rights education that is not grounded in reflexivity and solidarity with struggles for rights elides the transformative possibilities for authentic learning in HRE.

Increasing international and comparative discussions of human rights as an integral component of civic education necessitates greater attention to more robust methods—through portfolios, project-based learning, service learning, and civic engagement as curriculum—to authentically assess transformative forms of rights-based education. There is no doubt that there is value to human rights education – certainly Fatima and Zau would attest to this. But narrow indicators of causation and value cannot be superimposed upon larger mandates for civic engagement, citizenship competencies

and education for peace and human rights. To do so would be to limit the power and force of such holistic educational projects and to eclipse their vast cosmopolitan imaginaries.

References

- Bajaj, Monisha. (2012). *Schooling for Social Change: The Rise and Impact of Human Rights Education in India*. New York: Bloomsbury.
- Bajaj, Monisha, Canlas, Melissa, & Argenal, Amy. (2017). Between Rights and Realities: Human Rights Education for Immigrant and Refugee Youth in an Urban Public High School. *Anthropology & Education Quarterly*, 48(2).
- Bajaj, M., Cislighi, B. & Mackie, G. (2016). Advancing transformative human rights education. Appendix to G. Brown (Ed.) *The Universal Declaration of Human Rights in the 21st century: Report of the Global Citizenship Commission*. Cambridge, UK: Open Book Publishers. Available at <https://www.openbookpublishers.com/shopimages/The-UDHR-21st-C-AppendixD.pdf>
- Flowers, Nancy. (2003). *What is Human Rights Education? A Survey of Human Rights Education* Retrieved from http://www.hrea.org/erc/Library/curriculum_methodology/flowers03.pdf
- Meyer, J. W., Bromley-Martin, P., & Ramirez, F. O. (2010). Human Rights In Social Science Textbooks: Cross-National Analyses, 1970- 2008. *Sociology of Education*, 83(2), 111-134.
- Tibbitts, Felisa. (2002). Understanding What We Do: Emerging Models for Human Rights Education. *International Review of Education*, 48(3-4), 159-171.
- Varenne, Herve & McDermott, Ray. (1998). *Successful Failure: The school America builds*. Boulder, CO: Westview.

Are global learning metrics feasible?

Aaron Benavot, UNESCO Global Monitoring Report, University at Albany-SUNY

For decades global information on education mainly emphasized comparable data on access to and completion of school instead of what students take away from their schooling experience. Since 2000 there has been a pronounced shift to measuring learning outcomes, with more and more countries (and civil society organizations) assessing student learning in national, regional and international assessments. The desirability and feasibility of developing comparable measures of learning outcomes has thus become a global issue, vigorously debated among policy analysts, donors and researchers.

Proponents of cross-country comparable measures of learning outcomes argue they are long overdue. The plight of millions of poor and marginalized children, who fail to master basic competencies during primary school grades, while recognized, fails to garner sufficient policy attention and threatens the achievement of the ambitious sustainable development agenda. Global measures of learning would help to address this.

Global monitoring of learning outcomes would push governments to prioritize learning and ensure that all children acquire core knowledge and master basic skills. Comparable measures of learning can also promote a culture of transparency, and evidence-oriented policy making. They can contribute to public debates of desired learning outcomes, and improve international partnerships. They can also help countries develop their capacity for analysing results and assessing a wider range of skills and competencies.

Yet many have raised concerns over the nature and value of global measures of learning. The exclusive focus on basic or minimum proficiencies in reading and numeracy, which are more amenable to measurement, risks marginalizing the value of a wider range of subjects and competences and can weaken national curriculum priorities. Moreover, measuring learning only in terms of literacy and numeracy skills gives a slanted portrait of the breadth of outcomes that schooling enables children to acquire.

Cross-country measures of learning, whether intended or not, can lead to country rankings, whose value remains problematic. League tables on learning can discourage country participation in assessments that contribute to effective policy reform over time. Also, learning contexts are diverse, which makes it difficult to develop and interpret comparable measures of learning both in language and mathematics.

The costs of comparative learning assessments, which can be a substantial burden for poorer countries, are likely to result in requests for funding support from international aid agencies.

Finally, while large-scale assessments are useful in tracking system-level performance, evidence is limited on how useful they are in enhancing teacher training and classroom practices and in reducing inequalities in learning over time. The adage ‘don't value what you measure, measure what you value’ can possibly unite the international education community under the common cause of improving education quality. Student proficiencies in reading and mathematics, which represent key foundational skills, are highly valued. However, measuring proficiency levels in these areas requires sensitivity to national and linguistic contexts. Monitoring these and other learning outcomes should be an ‘open source’ project developed in a collaborative and transparent manner. Coordination will be needed to

efficiently allocate financial resources for improving the comparability of learning measures across diverse contexts. Robust reporting of learning globally depends on reaching maximal consensus on content, quality and process.

The value and purposes of tracking learning over time depend to a considerable extent on whose perspective is involved. For national policy makers the monitoring of learning gains or losses provides an indication of system performance. It can also serve as a marker of on-going reform efforts to improve teacher preparation, broaden the relevance of instructional materials and assess the effectiveness of interventions targeting underachieving students. Keeping tabs on student learning at the local school level can help principals compare the achievements of their students with those of similar backgrounds in the same district, region or province. Such yardsticks of student learning better situate the specific learning challenges that individual schools face.

However, carrying out robust and internationally comparable learning assessments involves an enormous commitment of time, effort and resources. As such, they are most cost effective when they serve multiple purposes, including but not limited to providing a platform for global monitoring. Moreover, if these assessments are meant to track progress over time, rather than serve as a one-off measure of student learning, they require sustainable financing and political commitment, independent of a particular government or minister in power. In many parts of the world, these conditions are difficult to obtain and solutions on the ground are inevitably imperfect.

Clearly the development of global measures of learning is not simply a technical issue, but also a political one. Opposition to learning assessments in some countries has become quite salient among parents, teacher associations and political parties, for example. To address these concerns the GEM Report recommends that governments embrace open and inclusive approaches that prioritize the needs and capacities of their countries based on the criteria of inclusivity, efficiency and feasibility. And they should make concerted efforts to build consensus around the content, quality and process of assessment activity.

Can Global Learning Metrics be Pedagogically Innovative?

Can global learning metrics be pedagogically innovative?

Radhika Gorur, Deakin University

Global learning metrics depend on standardisation, which is antithetical to innovation. Everything from the definition of ‘school’ and ‘student’, to what is to be valued as ‘outcomes’ of education, to how to measure ‘performance’, ‘quality’ and ‘equity’ must be standardised before any global comparative metrics can be developed. In projects of large-scale measurement, the outliers (and innovators, by definition, are non-standard) are regarded as a nuisance, because they skew the data and distort patterns. They are often weeded out in the analyses.

So from the outset, global learning metrics are projects of convergence and harmonisation, rather than of accommodating difference – and ‘difference’ is key to innovation. It is a normative and regulatory mechanism that pushes for conformity and suppresses the outliers, the idiosyncratic and, by extension, the innovators. By its very nature, the standardisation that underpins global metrics is designed to ignore rather than highlight innovation.

One of the most striking effects of global learning metrics over the last couple of decades has been that education policies globally have been in a constant state of ‘reform’. In many countries, rankings on international large-scale comparative assessments (ILSAs) have mobilised anxieties about ‘slipping’ and ‘falling behind’ in an assumed global, zero-sum game, and a slew of reforms have been mobilised in a bid to reverse this, or to ‘stay on top’ in the league tables. In many cases, such anxieties have led to more stringent accountability requirements and closer monitoring of student performance as well as of teachers and schools.

On the whole, I would argue, this has led to more conservative policies and schooling practices than it has to pedagogic innovation. To support this argument, I raise the following points:

- The standardisation of norms in school and teacher practice encouraged by global learning metrics tends to run counter to innovation and risk-taking. It encourages perverse effects through gaming and allows teachers to innovate only within narrowly-framed curriculum spaces
- Global learning metrics are aimed at policy and there is a big gap (of both logic and of practical politics) between policy reform and classroom practices
- Global learning metrics attempt to steer reforms to emulate ‘best practice’ – but ‘reform’ is not the same as ‘innovation’. Pedagogic innovation, in particular, is rarely prompted by imposed reforms. It is necessarily led by teachers and schools.

Policy Focus

The immediate focus of large-scale assessments and global metrics is to inform policy rather than classroom practice or pedagogy directly. IEA for example, seeks to ‘assist policymakers in identifying the relative strengths and weaknesses of their education systems’¹, and the OECD highlights PISA’s ‘policy orientation, with design and reporting methods determined by the needs of governments to draw policy lessons’ as a key feature of PISA. Both TIMSS and PISA, however, also imply that, directly or indirectly, there are lessons from their assessments for classroom practice and for

¹ http://www.iea.nl/about_us.html

pedagogic innovation. TIMSS makes greater claim for providing data that may more directly inform classroom practices, since their assessments are classroom-based and linked to curriculum – by testing 8th graders in intact classes, TIMSS results can be directly linked to particular teachers and schools and students. The TIMSS video studies tried to make the connection between classroom practices and performance more explicit, and to focus more on documenting different practices. PISA, on the other hand, tests by age, and includes 15-year-olds scattered across grade levels and taught by different teachers. Nevertheless, it claims that it can point to ‘best practice’ and produces a range of ‘lessons for’ reports with recommendations for policy that could impact classroom practices.

International assessments can demonstrate some broad-brush patterns in the performance of countries on the learning areas being assessed. Regular cycles of assessment can provide useful information with regard to trends in performance. However, such assessments cannot bear the enormous burden that is being placed on them to ‘tell’ policymakers what to do (Gorur, 2011). The psychometricians and statisticians involved in these studies, many of whom I have interviewed over several projects, appear bewildered that so much faith is placed on these measures. Nevertheless, organisations such as the OECD actively seek to influence reform, and nations do indeed base reforms on global comparisons. The influence of the OECD on national policies and practices has been well documented (Gorur, 2016; Hopkins, Pennock, Ritzen, Ahtaridou, & Zimmer, 2008). But, basing reforms on average performance can be wasteful and even damaging, as we found in the case of Australia (Gorur & Wu, 2014). Rather than focus on within-country differences, directing resources into particular areas of need, and conserving all that is good in the system, a panic about how we are ‘slipping’ in international rankings has led to much more centralised control with sweeping, rather than focused and targeted, changes, to the detriment of our students and some of our states and territories, particularly those with large Indigenous populations. Secondary analysis, which looks beyond the rankings to more perform more detailed analysis, has its own pitfalls and limitations too (Rutkowski, Gonzalez, Joncas, & von Davier, 2010).

Global learning metrics are a product of globalisation as well as active contributors to it. As such, they promote, deliberately or otherwise, a convergence of policy by producing role models to emulate, rather than expanding the policy imaginary or encouraging policy innovation by highlighting a range of models. Although different nations understand and take up the ‘policy lessons’ produced by global comparisons quite differently, global metrics have led to the creation of a globalised policy field (Lingard, Rawolle & Taylor, 2005) which influences and limits the possibilities of innovation.

Requirements of Pedagogic Innovation

What is ‘pedagogic innovation’, and what does it entail? As a teacher and a curriculum leader for 25 years in schools in Nigeria, Oman, India and Australia, I have been engaged in many reform and innovation efforts. A principal difference I found between ‘reform’ and ‘innovation’ in schools is that reform involved the planned imposition of a pre-determined model, whereas ‘innovation’ often arose unexpectedly and proceeded in ways that were not pre-planned. Generally, successful innovation starts from questions or dissatisfactions of some sort. In one case, a major school-wide curriculum innovation effort arose from my puzzlement over why the six-year-olds in my Grade 1 classroom could not solve a certain mathematics worded problem. Global metrics may provide the impetus for reforms – they may advocate that certain practices be emulated. They speak in terms of ‘lessons to be learned’ and ‘models’ to learn from. In other words, even the reforms that global metrics may advocate are about conformity rather than innovation. Global comparisons provide role models and ‘reference societies’ (Waldow, 2014), rather than an invitation for doing something new and radical.

Global metrics aim to highlight ‘best practices’. The argument is that if a nation performs well on international surveys, the practices they use must be ‘best practices’ which we can learn from and apply – even if in a modified way – in our own countries. But there is no reason to believe that just because certain practices are used in high-performing nation, those practices are ‘best practices’. For all we know, that country might have performed even better with a different set of practices. Moreover, if we looked beyond the countries that performed well, we might find that low-performing school systems also used same practices! So the premise of the ‘best practices’ argument needs to be questioned.

Innovations typically are – or at least start off as – small-scale and local. Seldom are untested ideas funded as large-scale projects. As such, they are unlikely to be picked up in ILSA surveys. So even where innovative practices might be occurring, global metrics are unlikely to pick them up in their sample-based surveys and so they are unlikely to highlight them. But, just to play along with the argument that global comparisons highlight ‘best practice’ for a moment, let us look at some of the high-performing nations and what kind of practices they use. If we examine the East Asian economies – Shanghai (China), Hong Kong (China), Singapore and Korea – there are a number of issues with the practices used there. Generalising broadly, a major motivator of good performance is the shame attached to poor performance. Families spend enormous amounts of money on cram schools, and children spend enormous time and energy in these schools. Surely these are not the practices that we would like to import into other countries.

Empowering Teachers

Pedagogic innovation is led by teachers and schools. However, in this era of accountability and audit, of which ILSAs are a part, there is great pressure on teachers to conform and little incentive to experiment. Although countries like Finland provide a model for employing highly qualified teachers and then trusting their professionalism and allowing them to teach with minimal surveillance, this is not the model that has been taken up globally. Rather, teacher accountability has been ratcheted up with impossible measurements such as what value individual teachers add to a student’s performance (Value Added Measures). There is a paradoxical valuing of teachers as the single biggest in-school influence on student performance, whilst at the same time assuming that they must be constantly monitored and that, unless they are held accountable and incentivised in various ways, they would tend not to do their best. In most countries, teachers are not included – or are included in tokenistic ways – in policy discussions and reform plans.

Although this kind of climate is not a direct result of global metrics, it is certainly one to which they have contributed. Every country wants to be in the top five in international rankings, as if that were a prize in itself (Gorur & Wu, 2015). Many countries have introduced national assessments and present the results in comparative formats which makes student performance (often linked to teacher performance) ‘high stakes’. This is not the kind of ethos that supports, let alone encourages, innovation.

References

- Gorur, R. (2011). ANT on the PISA Trail: Following the statistical pursuit of certainty. *Educational Philosophy and Theory*, 43(S1), 76-93.
- Gorur, R. (2016). Seeing like PISA: A Cautionary Tale about the Performativity of International Assessments. *European Educational Research Journal*, Online First.
doi:10.1177/1474904116658299

- Gorur, R., & Wu, M. (2015). Leaning too far? PISA, Policy and Australia's 'top five' ambitions. *Discourse-Studies in the Cultural Politics of Education*, 36(5), 647-664.
- Hopkins, D., Pennock, D., Ritzen, J., Ahtaridou, E., & Zimmer, K. (2008). External Evaluation of the Policy Impact of PISA (Vol. EDU/PISA/GB(2008)35/REV1). Paris: OECD.
- Lingard, R., Rawolle, S. & Taylor, S (2005). Globalising Policy Sociology in Education: Working with Bourdieu. *Journal of Education Policy* 20(6), 759-777
- Rutkowski, L., Gonzalez, E., Joncas, M., & von Davier, M. (2010). International Large-Scale Assessment Data: Issues in Secondary Analysis and Reporting. *Educational Researcher*, 39(2), 142-151.
- Waldow, F. (Producer). (2014, 2 June 2016). Projecting Images of the "Good" and the "Bad" School: Using Top Scorers in Large-scale Assessments as Reference Societies. [Keynote video]
Retrieved from <http://international-assessments.org/videos/>

Can global learning metrics be pedagogically innovative? Next big thing in education: Small data

Pasi Sahlberg and Jonathan Hasak*
Published in Washington Post, 9 May 2016

One thing that distinguishes schools in the U.S. from schools around the world is how data walls, which typically reflect standardized test results, decorate hallways and teacher lounges. Green, yellow, and red colors indicate levels of performance of students and classrooms. For serious reformers, this is the type of transparency that reveals more data about schools and is seen as part of the solution to how to conduct effective school improvement. These data sets, however, often don't spark insight about teaching and learning in classrooms; they are based on analytics and statistics, not on emotions and relationships that drive learning in schools. They also report outputs and outcomes, not the impacts of learning on the lives and minds of learners.

After The No Child Left Behind Act became law in 2001, education legislation in the U.S. required all students in grades 3 to 8 each year and once in high school to be tested in reading and mathematics using external standardized tests. On top of that states had their own testing requirements to hold schools and teachers accountable. As a result, various teacher evaluation procedures emerged in response to data from these tests. Yet for all of these good intentions, there is now more data available than can reasonably be consumed and yet there has been no significant improvement in outcomes.

If you are a leader of any modern education system, you probably care a lot about collecting, analyzing, storing, and communicating massive amounts of information about your schools, teachers, and students based on these data sets. This information is "Big Data," a term that first appeared around 2000, which refers to data sets that are so large and complex that processing them by conventional data processing applications isn't possible. Two decades ago the type of data education management systems processed were input factors of education system, such as student enrolments, teacher characteristics, or education expenditures handled by education department's statistical officer. Today, however, Big Data covers range of indicators about teaching and learning processes, and increasingly reports on student achievement trends over time.

Despite the outpouring of data, international organizations continue to build regional and global data banks. Whether it's the United Nations, the World Bank, the European Commission, or the OECD today's international reformers are collecting and handling more data about human development than before. Beyond government agencies, there are global education and consulting enterprises like Pearson and McKinsey that see business opportunities in Big Data markets.

Among the best known today is the OECD's Programme for International Student Assessment (PISA) which measures reading, mathematical, and scientific literacy of 15-year-olds around the world. OECD now also administers an Education GPS, or a global positioning system, that aims to tell policymakers where their education systems place in a global grid and how to move to desired destinations. OECD has clearly become a world leader in the Big Data movement in education.

* Jonathan Hasak, based in Boston, is working to change public policies to better support youth who are disconnected from the labor market and disengaged from school. Follow him on twitter @JonathanHasak

Despite all this new information and benefits that come with it, there are clear handicaps in how Big Data has been used in education reforms. In fact, pundits and policymakers often forget that Big Data, at best, only reveals correlations between variables in education, not causality. As any introduction to statistics course will tell you, correlation does not imply causation. Data from PISA, for example, suggests that the “highest performing education systems are those that combine quality with equity.” What we need to keep in mind is that this statement expresses that student achievement (quality) and equity (strength of the relationship between student achievement and family background) of these outcomes in education systems happens at the same time. It doesn’t mean, however, that one variable would cause the other. Correlation is a valuable part of evidence in education policymaking but it must be proved to be real and then all possible causative relationships must be carefully explored.

The problem is that education policymakers around the world are now reforming their education systems through correlations based on Big Data from their own national student assessments systems and international education data bases without adequately understanding the details that make a difference in schools. A doctoral thesis in the University of Cambridge, for example, recently concluded that most OECD countries that take part in the PISA survey have made changes in their education policies based primarily on PISA data in order to improve their performance in future PISA tests. But are changes based on Big Data really well suited for improving teaching and learning in schools and classrooms?

We believe that it is becoming evident that Big Data alone won’t be able to fix education systems. Decision-makers need to gain a better understanding of what good teaching is and how it leads to better learning in schools. This is where information about details, relationships and narratives in schools become important. These are what Martin Lindstrom calls Small Data: small clues that uncover huge trends. In education, these small clues are often hidden in the invisible fabric of schools. Understanding this fabric must become a priority for improving education.

To be sure, there is not one right way to gather Small Data in education. Perhaps the most important next step is to realize the limitations of current big data-driven policies and practices. Too strong reliance on externally collected data may be misleading in policy-making. This is an example of what small data look like in practice:

1. Reduced census-based national student assessments to the necessary minimum and transfer saved resources to enhance the quality of formative assessments in schools and teacher education on other alternative assessment methods. Evidence shows that formative and other school-based assessments are much more likely to improve quality of education than conventional standardized tests
2. Strengthened collective autonomy of schools by giving teachers more independence from bureaucracy and investing in teamwork in schools. This would enhance social capital that is proved to be critical aspects of building trust within education and enhancing student learning.
3. Empowered students by involving them in assessing and reflecting their own learning and then incorporating that information into collective human judgment about teaching and learning (supported by national big data). Because there are different ways students can be smart in schools, no one way of measuring student achievement will reveal success. Students’ voices about their own growth may be those tiny clues that can uncover important trends of improving learning.
4. Edwards Deming once said that, “without data you are another person with an opinion.” But Deming couldn’t have imagined the size and speed of data systems we have today. Automation that relies on continuously gathered data is now changing our daily lives. Drivers today don’t

need to know how to use maps anymore when they can use smart navigators that find them the best routes: airline pilots spend more time flying on autopilot than by hand. Similar trends are happening in education systems with countless reformers trying to “disrupt” schools as they are.

Big Data has certainly proved useful for global education reform by informing us about correlations that occurred in the past. But to improve teaching and learning, it behooves reformers to pay more attention to small data – to the diversity and beauty that exists in every classroom – and the causation they reveal in the present. If we don’t start leading through small data we might find out soon enough that we are being led by big data and spurious correlations.

Can global learning metrics be pedagogically innovative? Educational Prosperity: An assessment framework for global metrics that are pedagogically innovative

J. Douglas Willms, University of New Brunswick and The Learning Bar Inc.

The capacity of a society to develop young peoples' literacy skills and well-being depends on its ability to provide the right kinds of human and material resources to support healthy child development from conception to adolescence, and beyond. Educational Prosperity is an assessment and reporting framework that can be used to evaluate the capacity of a school district, state, or country to develop children's literacy skills and well-being, to set goals for increasing a school system's capacity, and to monitor progress towards meeting those goals. The framework follows a life-course approach, with key outcomes for each of six stages of development. These outcomes are referred to as 'prosperity outcomes'. The family, institutional and community factors that drive these outcomes are called 'foundations for success'. The underlying premise is that if countries build strong foundations for success for each stage of children's development, its children will thrive.

In most evaluation systems, the links between the construction of measures, the methods used to collect data, the analytics, and the reporting strategies are 'loosely coupled'; that is, the elements are disconnected such that they have little impact on school policy and classroom practice.

First, and perhaps the biggest 'disconnect,' is that the metrics lack relevance for classroom practice. For example, the Programme for International Student Assessment (PISA) and the Trends in International Mathematics and Science Study (TIMSS), and many state and national monitoring systems, collect data on students' reading, mathematics, and science skills. However, teachers usually teach subjects based on a fairly prescriptive curriculum. Language arts teachers, for example, do not explicitly teach reading skills.

Second, global learning metrics typically cover a narrow range of skills. National tests targeted to a particular grade level typically have test items for that grade level, and a few items at the adjacent grade levels. However, by the end of grade 3, the range of children's reading and mathematics skills usually span at least four grade levels, and by the end of grade 8, they can span as much as eight grade levels. The distributions tend to be negatively skewed, with a large percentage of students scoring at the low end of the distribution. Many of these students have moderate to severe learning needs and are several years behind their peers in their development of basic skills. Most global learning metrics do not adequately capture the abilities of students with low skill levels.

Third, some critics argue that global learning metrics do not capture higher-order thinking skills. While this may be the case with some assessments, there have been considerable improvements in assessment design over the past twenty years. Many of these improvements are owing to the work of researchers who developed tests for large international studies such as PISA and TIMSS.

Fourth, some tests used for global metrics are culturally biased, yielding results that do not allow for fair cross-national comparisons. This is also a concern for assessments designed to measure the skills of students across language and cultural groups within countries.

The fifth 'disconnect' is that cross-sectional assessments are often used to hold school principals and teachers accountable. The contributions of school staff – the so-called 'value added' – cannot be adequately assessed with cross-sectional data.

The analysis and reporting strategy for Educational Prosperity has three explicit links to national and local policy and practice. First, it allows countries to align data collection with goals at all levels within the system. Second, the data collected within this framework have immediate implications for educational policies involving decisions about the allocation of resources or the assessment of interventions aimed at changing long-standing structural features of schools. Third, the approach enables countries to set targets that are consistent with state or national goals, or in the case of low- and middle income countries, to set targets consistent with the United Nations post-2015 framework.

The Learning Bar has developed two assessments that are being used in national and provincial planning to measure prosperity outcomes and the foundations for success. These assessments not only provide global learning metrics, they are also designed for improving teaching and learning. In both cases, the data collection cycle is short and reports are provided immediately upon completion of data collection. The emphasis is on collecting data to provide leading indicators rather than trailing indicators. Leading indicators focus on the learning and teaching needs of students and teachers, looking forward, rather than trailing indicators, which are focused on past performance and issues concerning accountability.

One of these assessments is the Early Years Evaluation (EYE), which consists of two complementary assessment tools that help educators monitor the overall development of children as they prepare for and make the transition to school. They assess skills in five domains consistent with the framework for school readiness developed by the US National Education Goals Panel and currently used by UNICEF. The EYE-Direct Assessment (EYE-DA) is designed to measure the developmental outcomes of children aged 3-5 years using an engaging and play-based format of assessment. The EYETeacher Assessment (EYE-TA) provides a systematic framework that kindergarten teachers can use to structure their frequent observations and informal assessments. Results from both assessments are provided immediately.

The EYE assists teachers by providing formative, instructionally relevant information they can use to plan their day-to-day learning activities. Meeting this need has been the primary focus of the assessment. The reports are often used alongside other assessments to identify students who are encountering difficulty and may need extra support. Teachers use the results to involve parents in meaningful ways, such as providing suggestions of activities that parents can do at home to strengthen their children's skills.

The second assessment that provides leading indicators is the OurSCHOOL evaluation system, which includes student, teacher and parent surveys that provide reliable data on factors known to affect student outcomes. At the school district level, a completely customized survey can be created by choosing from bank of over 50 measures. Many of these measures are metrics for the foundations for success at various levels of the Educational Prosperity model. In some jurisdictions, a senior educator 'owns' the measures for each foundation. The person who owns 'quality instruction', for example, is tasked with ensuring that stakeholders at all levels of the system – the Minister and senior staff, district superintendents, school principals, teachers, parents, and students – understand how quality instruction is being measured, the annual targets for improvement, and various strategies for improvement. The Ministry and the school districts direct their attention to providing the necessary professional development and support materials for school improvement. When the system is working well, the emphasis at the school level is on using data to build a strong foundation for student success.

Can Global Learning Metrics be Culturally Responsive?

Can global learning metrics be culturally responsive?

Supriya Baily, George Mason University

One of the most popular myths of the 21st century is that school can transform your life. No matter your socio-economic status, your resident status in a country, your private role as a parent or an employer, or your participation in global policy forums, there is often an unquestioning belief in the power of schools and the role of education in peoples' lives. Moving towards a focus on learning from simply accessing education has been welcome shift, as evidence in recent decades shows that enrollment, access and physical presence in school do not simply promote learning. But while taskforces, researchers, policy makers and popular media promote the greater visibility of learning as an end result, critical voices are starting to ask more loudly "Learning towards what end?"

In the description of this session, the program asks whether the "concept of global learning metrics is based on the assumption that there is an agreement about what constitutes "good" and "quality" education worldwide" while also accepting the fact that these metrics have often "neglected the diversity of cultural contexts and educational systems." The program also asks whether there is a global "core of fundamental knowledge, skills and competencies that are relevant across different countries?" and if metrics can "capture the dynamics" of the intersectionality of student identity while also being more culturally responsive in light of the irrefutable fact that there are uneven power relations around the world.

The simple answer to those questions might be "No – global learning metrics cannot be culturally responsive in light global inequities and the dynamics of difference". Such an answer then would ignore the critical importance of ensuring school is a successful place for students, and that learning across multiple domains ought to occur (Learning Metrics Taskforce, 2013). While the symposium might not focus on defining what a "good", "quality" and "successful" education might mean, I do approach this conversation from the perspective that there is such a thing as a good education, but that the terms are relative, contextual and inequitable in delivery and outcomes. The larger argument I am making centers on two issues surrounding measuring learning. The marginalization of these two areas in the conversation are relevant to why global learning metrics might flounder in light of equity and responsiveness.

The first centers on how the choice of what, how, when and where to learn is driven by a single global ideology which by its sheer force and influence is incapable of providing space for any sort of culturally responsive measurement of true global diversity and agency. The influence of a neoliberal ideology continues to ensure that the expectations and outcomes of education remain central to the needs of the workplace. Even spaces that previously were immune or resisted the push of neoliberal ideologies are struggling (see Pearlstein, 2016). The greater focus on the knowledge economy and the ensuring efficiency of educational markets make it more challenging to address the needs of those on the margins of equity in education (Rizvi & Engel, 2009).

The intersection of colonialism, patriarchy and neoliberalism has created metropolises (Connell, 2010) that promote the ideas of equitable systems and structures while maintain the structure and systems of oppression. Connell (2010) argues that school systems "were constructed in the colonies at much the same time as they were being developed in the metropole" and while "the periphery becomes a source

of data, and a site of application...the concentration of data and the moment of theorising occurs in the metropole...Therefore intellectuals living in the periphery are strongly oriented to the metropole” (p. 608). So while the exercise of developing equitable learning metrics and ensuring the views of a wide variety of stakeholders are embedded in the outcomes and opportunities, the metrics to be followed are often viewed or adapted through the lens of what we are already conditioned to see.

The second issue focuses on the role of teacher education and the preparation of teachers in the measurement of learning especially across increasingly diverse student bodies. The Taskforce on Global Learning Metrics suggests that “globally tracked indicators should be aligned with what is measured nationally and in schools or classrooms, while measurement at the national level should be aligned with the competencies measured in classrooms or schools” (p. 3). The report suggests that teachers ought to be the first line of measurement but that such evaluation is often summative and administered in the form of final exams (LMTF, 2013). Student achievement is linked to culturally responsive teaching (Liebtag, 2016) but yet, culturally responsive teaching itself is not as widespread and few teacher education programs excel with such types of courses in their teacher preparation programs. Teacher assessment and analysis of student learning “needs to vary, in relation to the learners, the context and the subject matter knowledge” (Richert, Donahue & LaBoskey, 2009, p. 648). Without a focus on ensuring culturally relevant pedagogy that takes into account all forms of knowledge in a community, it seems impossible for learning to be measured at the point of instruction in a culturally responsive manner.

If we decide that learning metrics are desirable, feasible and pedagogically innovative, we have to ask if they are culturally responsive. Issues of culture are integrally linked to power. As long as the ideology of education is buttressed by a neoliberal stance, and teachers are judged by the assessments they have to scale, the ability for metrics to be culturally responsive seems to be aligned to rhetoric than practice.

References

- Connell, R. (2010). Kartini's children: on the need for thinking gender and education together on a world scale. *Gender and Education*, 22:6, 603-615.
- LMTF (Learning Metrics Task Force). (2013). *Toward Universal Learning: A Global Framework for Measuring Learning*. Report No. 2 of the Learning Metrics Task Force. Montreal and Washington: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.
- Levine, A. (2010). Teacher education must respond to changes in America. *Phi Delta Kappan*. October 2010.
- Liebtag, E. (2016) Culturally responsive pedagogy. Retrieved from <http://www.globalteachereducation.org/culturallyresponsive-pedagogy>
- National Education Association. (2010). Today's teacher issues: Latest statistics on teachers. Retrieved on September 15th 2010 from http://educationalissues.suite101.com/article.cfm/nea__todays_teacher_issues#ixzz0r7uTPp00
- Pearlstein, S. (2016, September 16). Meet the parents who won't let their children study literature. *The Washington Post*. Retrieved from <https://www.washingtonpost.com/>
- Rizvi, F. & Engel, L. C. (2009). Neo-liberal globalization, educational policy, and the struggle for social justice. In W. Ayers, T. Quinn & D. Stovall (Eds.), *The handbook of social justice in education* (pp. 529-541). Rowman & Littlefield Publishers.
- UNESCO (2016). Education statistics: Percentage of female teachers by education level. Retrieved from <http://data.uis.unesco.org/index.aspx?queryid=178&lang=en>

Can global learning metrics be culturally responsive?

Stafford Hood, University of Illinois at Urbana-Champaign

I am largely in agreement that the questions being raised at the symposium about global learning metrics and approaches to assessment are indeed some of the “right” questions. I have come to accept the personal reality that one cannot always be expected to have the right answer because it is too often the case that our reasonably good answers typically only raise more questions. As I think it should be. However, it is imperative that we have the right questions. I concur that the question as to whether global learning metrics can be culturally responsive, (within the context of hopefully some consensual agreement about what constitutes good and quality education) is one of the right questions. My initial reflex answer to this question is: I hope so; it depends; and why should it?

I was pleased that the conference organizers realistically acknowledged that any consideration about global learning metrics must accept the pervasive and long standing pattern to ignore/neglect any meaningful consideration of culture and cultural contexts that continues to exist in too many places. With this in mind, the expectation for any reasonably acceptable answer about whether global learning metrics can be culturally responsive has multiple layers of complexity, least of all being what different cultural groups believe should be good or quality education-- that is in their best interest. At the same time, of equal importance is what each of these groups may consider to be acceptable evidence that their children have received a good or quality education-- that is in their best interest.

I continue to believe that the assessment tools and systems we use should be responsive to the cultural background and experiences of our traditionally disenfranchised racial minorities, even if some are culturally specific. I also believe that the assessment devices and systems we use should be culturally valid to increase the possibility that they provide accurate and useful information to effectively educate these students. Validity is at the core of any consideration about assessment instruments we use to measure student learning. However, this simplicity becomes a bit more complex when the question is whether the instrument is measuring student learning equally well for all students, particularly those who have been traditionally disenfranchised, by their educational, economic, social and political systems. This is where it starts to get interesting.

It is the case that large and historical patterns of differences in the test performance of groups based on race, gender, and income have been too often the rule rather than the exception. Therefore, forcing one to wonder how much can we trust the accuracy and legitimacy of interpretations based on these scores. For example, the historical disproportionately low performance of African American students on standardized achievement tests, when compared to their white counterparts, continues to be prevalent in our conversations. This pattern makes it difficult to easily dismiss Keena Arbuthnot's (2014) question, as to whether there is “...some issue surrounding Black students or their culture that can explain why the tests could possibly be measuring something different for them in comparison to White students”?

My early thinking on culturally responsive assessment in the late 1990's was informed by the clarity that was emerging from the work of Gloria Ladson-Billings, Jaqueline Jordan Irvine, Carol Lee, and others around culturally relevant/responsive pedagogy. I pondered, with a couple of others about how to intervene in the test development and validation process through statistical and judgmental bias review approaches. I argued for the development of assessments that:

“...are grounded in the cultural contexts of examinees of color. If it is true that culturally responsive forms of assessment can quite possibly improve the assessment of what examinees of color know and can do relative to specific learning outcomes, then it is incumbent upon us to begin discussing how these devices can be developed and validated” (Hood, 1998 p. 188).

Solano-Flores and Nelson-Barber (2001) provided a clearer and more substantive articulation on cultural validity in assessment. They defined cultural validity as:

“...the effectiveness with which...assessment addresses the socio-cultural influences that shape student thinking and the ways in which students make sense of ...items and respond to them. These socio-cultural influences include the sets of values, beliefs, experiences, communication patterns, teaching and learning styles, and epistemologies inherent in the students’ cultural backgrounds, and the socioeconomic conditions prevailing in their cultural groups” (as cited in Solano-Flores 2011, p.3)

Solano-Flores (2011) would later remind us that any product produced by a human being will be imperfect. Therefore tests are also imperfect and are in fact “cultural artifacts” reflecting the dominant culture that uses them for inclusion and exclusions to societal benefits. He asserted that tests are:

“...created with the intent to meet certain social needs or to comply with the mandates and legislation established in a society; they are written in a language (dialect of that language) used by those who develop them; their content is a reflection of the skills, competencies, forms of knowledge, and communication styles valued by a society—or the influential groups of that society;

He would further state that tests “... assume among test takers full familiarity with the contexts used to frame problems, the ways in which questions are worded and the expected ways to answer those questions.”

As I maintain a certain level of optimism, it remains difficult not to hear the contrarian voices. For example, Mehan (2008) rightfully reminds us that students from poor, ethnic, and linguistic minority backgrounds are restrained in the starting blocks before they allowed to run the “race for success”. Whether this is by accident or design I leave for each of us to individually decide. Still I think we can generally agree that in too many of our classrooms, racial minority and poor students find themselves in situations that are “toxic or disabling” with the prospect of learning a distant possibility with assessment seemingly being used “...for purposes of behavioral control and for coercion to learn” (Boykin, 2013). Believing this to be the case, I agree with Boykin that the success of our traditionally underserved students in the classroom and their performance on high stakes tests is a social justice issue. He states:

“In all, the social justice argument is that the bias, problematics or illegitimacy of such measures must be exposed and rectified since in their present forms these assessments do not allow the abilities and propensities of certain groups to be accurately or adequately ascertained... (p.33)

He goes on further to say:

“we need to determine how such assessments can more truly discern the performance capabilities of certain groups without undermining the psychometric quality of such assessments; without unduly penalizing those who fare well on such measures in their present forms; and without inappropriately undermining the pursuit of excellence in the standards of the outcomes sought” (ibid.)

Indeed, we must build better assessments that allow for the abilities of “certain groups” to be accurately measured while also building these instruments so that the psychometric quality is not compromised. This will not be an easy task but it certainly must be done.

Acknowledgements

The Center for Advanced Studies in Global Education and edXchange would like to thank its partners and sponsors for their support of the Inaugural Symposium of the Comparative and International Education Society. Thank you to Open Society Foundations, the Comparative and International Education Society, and ASU International Development for your generous support.

The Center for Advanced Studies in Global Education

P.O. Box 871611, Tempe, AZ 85287-1611

CASGE@asu.edu

(480) 727-5346

education.asu.edu/casge



@asucasge



ARIZONA STATE UNIVERSITY

