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 systems 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 rag 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 Educacion para América Latina y el Canibe 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 Barett 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 get displaced 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 advise 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 simplification 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.