Measurement & Evaluation

Cracking the Code on Social Impact

An emerging measurement approach emphasizes the value in standardizing outcomes for the whole social sector.

By Jason Saul | Feb. 6, 2014

Over the past few decades, practitioners, evaluators, and academics have struggled to organize, measure, and understand social change. We have made a number of important advances, including more rigorous control studies, digitization of 990 data, outcomes tracking software, and improved reporting. One important challenge evaluators have faced in the social sector is standardization: How can we learn from past efforts if we cannot systematically compare one socially focused program to another? Researchers have tried to solve the “apples to oranges” problem in a number of ways. In the 1980s, the Urban Institute’s National Center on Charitable Statistics (http://nccs.urban.org/) (NCCS) created a common code for classifying nonprofit organizations by entity type (http://nccs.urban.org/classification/NTEE.cfm), and later created another system to classify program services and beneficiaries (http://nccs.urban.org/classification/NPC.cfm). Others have tried to standardize performance metrics using “shared measurement systems” such as IRIS (http://iris.thegiin.org/) and the Cultural Data Project. (http://www.culturaldata.org)

Still, these efforts fall short of codifying the true results of an organization’s programmatic efforts: outcomes.

While much of the focus on outcomes has centered on trying to measure them idiosyncratically, one organization at a time, I believe—and our work at Mission Measurement (http://missionmeasurement.com/) is demonstrating—that the real value is in standardizing the use of outcomes for the whole sector. Here’s why this approach works:

- First, standardizing outcomes enables us to organize and identify social programs in a more meaningful way. Instead of researching programs based on subject area (for example, education, youth development, or the arts), we can base research on the benefits programs aim to produce (for
example, improving college readiness, increasing access to public services and supports, or encouraging artistic expression).

- Second, common outcomes create a universal common denominator for benchmarking and comparison—something the sector has long sought. So we can now apply measures such as cost per outcome and social return on investment, and we can do this on an ever-increasing scale, as we standardize and universally adopt efficacy rates.

- Finally, program design and learning is much more efficient. Researchers can more easily identify program elements that increase efficacy and cross-sector synergies that conspire to produce outcomes.

My thinking about this approach to measurement—what I call “universal outcomes taxonomy”—dates back to my 2004 book *Benchmarking for Nonprofits* (http://www.amazon.com/Benchmarking-Nonprofits-Measure-Improve-Performance/dp/0940069431) and a subsequent effort I led at the Urban Institute and the Center for What Works to create a prototype for classifying social outcomes (http://www.urban.org/UploadedPDF/411404_Nonprofit_Performance.pdf). Since then, we have carefully documented more than 78,000 outcome data points (78,369 to be exact) from more than 5,800 social programs. Two years ago, we focused a team of researchers on the daunting task of systematically cataloguing those outcomes—removing duplicates, standardizing language, creating hierarchies, and developing a universal taxonomy. Not surprisingly, we found that many of the outcomes were the same, though articulated differently. For example, one organization might state its objective as “student achievement,” another as “academic achievement,” and another as “improving test scores.” In all, we identified 132 common outcomes across the entire social sector. We then indexed these outcomes by program type and sub-type, and classified them into a functional taxonomy (see below). We also added geo-coding and beneficiary codes to better contextualize outcomes.

But while it is becoming clearer what value-seeking funders want, the problem remains that we still face a number of barriers to producing these data.

Today, a growing number of funders across the country are using this framework to organize, or “tag,” their grants and programs, making it possible to analyze their work on a portfolio level. They can measure the resources they have allocated to each outcome and the relative contribution of each program to a particular outcome, and aggregate the overall performance of the portfolio.
We believe that this new language holds great promise for the sector, and we have much work to do. First, we must enable widespread adoption of the taxonomy—by both funders and social service providers. Second, we need to build the capacity of organizations to select the right outcomes (what we call “sizing the outcomes”). And finally, we must continually curate and improve the taxonomy, with feedback from practitioners and researchers.

Jason Saul, founder and CEO of Mission Measurement (@MissionMeasure) is one of the nation's leading experts on measuring social impact. He is a faculty member at Northwestern's Kellogg School of Management, where he teaches social innovation and measurement.

If you like this article enough to print it, be sure to subscribe to SSIR!

Copyright © 2017 Stanford University. Designed by Arsenal, developed by Hop Studios
The Problem With The Impact Genome Project

Would the Impact Genome Project have predicted the impact of Martin Luther King?

By Emily E. Arnold-Fernández | May 29, 2014

The Impact Genome Project (http://www.ssireview.org/blog/entry/introducing_the_impact_genome_project) plans to use "big data and genomic analysis to measure, predict, and improve the outcomes of social programs."

It's a worthy goal: All of us involved in social change would love the power to predict outcomes, particularly when we're on the hook to funders for those results.

What may get lost, however, are the programs that matter most in the long run.

Sure, we can use data from prior bed-net distribution campaigns to predict whether a new bed-net campaign will successfully increase the number of bed net users (and thus decrease instances of malaria). We may even be able to use data from bed-net distribution campaigns to predict, albeit with less certainty, the outcome of a condom distribution campaign. Given how much government (http://www.ssireview.org/tags/Government) and philanthropic funding goes toward aid programs, that's important information.

But such programs represent only a small subset of the changes our world needs. Sometimes, yes, we need more people to use bed nets. Other times, we need laws and courts that protect journalists, appropriate police investigation and prosecution of gender-based crimes, an adequate system of contract enforcement so entrepreneurs can flourish, or an end to the internment of refugees in camps—a massive waste of human capital and a blatant violation of refugees' rights to work.

Can the Impact Genome Project predict the success of initiatives designed to achieve these ends? Would the Impact Genome Project have predicted Dr. Martin Luther King Jr.'s success or Harvey Milk's influence on marriage equality?
Even with bed nets, we probably don't see distribution by aid agencies as a sustainable model, long-term. Instead, we want policies that make bed nets a local imperative: government incentives for local production of bed nets, public health advertising encouraging bed net use, and a strong science curriculum in schools so that students—the next generation of parents, citizens, and leaders—understand why bed nets matter.

Even for bed nets, we ultimately want a new policy framework. Without it, social change is unsustainable.

If the Impact Genome Project can accurately predict programs that yield policy change, anywhere in the world, I'll be thrilled. I'll also be surprised. The factors that influence success are highly varied and often have little or nothing to do with the policy in question.

My organization's Thai affiliate, Asylum Access Thailand, for example, began work in 2010 aimed at getting a historic temporary asylum law on the books in Thailand. If passed, the law—technically an amendment to Thailand's existing Immigration Act—would mark the first time Thailand granted any legal status to refugees. Could the Impact Genome Project have predicted that, just as the draft law was ready for introduction to Parliament in 2014, one party would boycott elections and leave the country without a functioning government? Or could it have predicted that Asylum Access Tanzania's 2012 letter to the UN Special Rapporteur on Migrants, which initially soured our relationship with the Tanzanian government, would ultimately catalyze improvements in the government's approach to refugees when a new Minister of Home Affairs was appointed?

More likely (in my admittedly non-data-driven view), the Impact Genome Project will reinforce many funders' preferences for programs that yield immediate and easily measurable returns.

The Impact Genome Project's creators, Jason Saul and Nolan Gasser, liken their efforts to those used in the financial world. They're right—and that's the problem. Lenders prefer those whose credit-worthiness is easily measurable—even those with a history of nonpayment—over foreigners, young people, and others whose credit-worthiness may be excellent but whose credit history is unknown. Financial markets prefer companies with recurring quarterly profits over those that incur losses as they build a foundation for greater gains long-term. The Impact Genome Project similarly incentivizes a focus on short-term results.

At the Skoll World Forum last month, Saul said he hopes that the Impact Genome Project will lead to a
"social capital market," where outcomes are bought and sold. The problem is, the most sustainable and effective social change won't sell well.

If we're serious about effective social change, we need more incentives for funders to support efforts to shift policy frameworks. By developing and promoting the Impact Genome Project, Saul and Gasser create a countervailing incentive, pushing funders toward direct aid models. And by pushing for a social capital market, they replicate, rather than correct, the inefficiencies of existing capital markets.

Emily E. Arnold-Fernandez (@EDasylumaccess) is executive director of Asylum Access, an innovative organization dedicated to ending the widespread violation of refugees' human rights in Africa, Asia, and Latin America. She is also a visiting practitioner with Stanford's Freeman Spogli Institute and a former social entrepreneur in residence (SEER) with the Center for Democracy, Development, and Rule of Law.

If you like this article enough to print it, be sure to subscribe to SSIR!

Copyright © 2017 Stanford University. Designed by Arsenal, developed by Hop Studios
Appreciate the responses. It’s exciting to see all the synchronous thinking in the field. A few clarifying points about our research and its implications:

• One vs. Many. There are a number of quality organizations keeping lists of commonly used outcomes, metrics and other data points. For example, some of the sources other commentators have referenced offer great resource lists. In a 1.0 world this is quite valuable as we need to seed the field with new ideas and compile reference lists for practitioners to access. The value of creating a universal taxonomy, however, is to bring the field to some level of standardization. To achieve standardization, we as a field need to adopt a single, universal, definitive, unbiased catalogue of outcomes by which to organize our work. And that catalogue must to be continuously curated, normalized and made useful to funders and practitioners. To that end, we will be making the Outcomes Taxonomy available to funders through a variety of channels and intermediaries beginning in Q2 of this year.

• Outcomes v. Metrics v. Activities. We often tend to conflate these distinctive concepts, and I think this makes standardization even more elusive. The purpose of our research was to curate-out those differences and develop a taxonomy that was purely focused on outcomes. Outcomes are directional changes in social condition. Metrics or indicators are easily conflated with outcomes, but are not the same. Things like “level of physical mobility” “time spent in hospital” or “number of people trained” are metrics.

• Outcomes v. General Conditions. To be universal, and standardizable, the outcomes included in a taxonomy must be both discrete and measurable. Oftentimes practitioners will try to phrase general conditions like “Improvements in policy and legislation” or “economic performance of a local area” as outcomes. But these conditions tend to be vague (e.g. not discrete) and open-ended (e.g. non-measurable) and therefore not well placed within a taxonomy.

• Agnostic to Theory of Change. A taxonomy is simply a list of outcomes. The judgment comes in the user’s selection or valuation of one outcome versus another. Whether you value “emotional well-being” or “test-scores” as a factor in driving up graduation rates, the point is that we are all aiming at the same shared “outcome” of improving graduation rates. The taxonomy doesn’t relate to the way in which you intend to produce that outcome.

• Outcomes v. Innovation. People often conflate “outcomes thinking” with measurement. These are very different things. Simply clarifying the desired benefits of an intervention does not quash or inhibit an organization’s ability to innovate strategies that produce that benefit. In fact, quite the contrary – it opens the door to a wide range of strategies because we are simply controlling for the outcome, not the theory of change, the metrics or the type of intervention.