



How Do We Know It When We See It?

by Sally Leiderman

People engaged in racial justice work face considerable pressure to provide evidence that their organization's particular approach makes a tangible difference in people's lives. The pressure comes from their own sense of urgency, from their constituents and from funders. Many people who fund this work are under similar pressure, sharing that sense of urgency, having to account for their decisions, and, like practitioners, wanting to structure future decisions based on evidence that the work is creating improvements. Evaluation sits right at the nexus of these similar and sometimes competing pressures.

While many organizations are working explicitly to reduce the historical or contemporary consequences of structural racism (even if they don't call it that), any group working on, for example, improving housing, education or health and well-being of children, youth, families or older adults in the U.S. is working on reducing structural racism or its impacts. That is because they must find ways to acknowledge, diagnose and redress inequitable outcomes by race as part of that work. And because outcomes in these areas are influenced by multiple system and institutional policies and practices and by cultural ideas about what is "normal" or "valued," these groups are always working on reducing structural racism or its impacts, named or not.

How might these organizations and their funders measure progress, particularly before large-scale population changes are evident? As a field, can we begin to develop evaluation measures based on more accurate understanding of what it takes to reduce structural racism's impact across groups and issues? Are there common shorter term changes that predict long-term success at improving life expectancy for all racial groups? Increasing educational achievement for all racial groups? Wealth accumulation for all racial groups? Are there common steps or markers of progress across these topical areas or systems? How would we know them when we see them?

Evaluating Work With Racialized Goals

Evaluations of work undertaken to address structural racism often examine time periods when the goal has not yet been attained

(for example – rates of incarceration by race haven't changed substantially, rates of entry and graduation from college by race/ethnicity remain wide apart). So the evaluation is looking at work "in the middle," trying to draw conclusions about whether work in the short run is going to make any difference in the long run. To see how tough this is, it's useful to ask, "If I were evaluating the civil rights movement, what conclusions might I have drawn about its long-term effectiveness in 1958? In 1964? Now?"

In addition, there is little consensus among practitioners, funders and evaluators about what progress towards reduced structural racism looks like, making evaluation efforts particularly challenging. For example, evaluation that focuses on "reducing the achievement gap" as an issue of teacher training and student preparation only, without considering the allocation of resources across schools or the influence of public violence on cognitive development of children in persistently under-policed neighborhoods, may contribute to setting unreasonable expectations or fail to capture important interim successes.

Logic Models and Theories of Change

Logic models lay out a measurable set of short-, intermediate- and long-term outcomes to which a group wants to hold itself accountable. Social scientists developed this tool specifically to help work out a negotiated understanding of what results are important to look at in that "middle" stage before the big, obvious changes can be observed, and as a process for building consensus on the combination of factors to be considered.

Theories of change are most often pictures that describe how the various parts of an effort are expected to contribute to the outcomes specified in the logic model. Like logic models, they provide an opportunity for all parties to understand each other's perspectives and worldviews about how change happens, what success looks like and expectations about reasonable accomplishments. The more the assumptions that underlie these issues surface (see sidebar) and are discussed, the more useful these tools will be for planning and evaluation.

The utility of these tools for evaluation of efforts with racially specific goals depends in large part on the extent to which all parties have, and are willing to apply, a racialized lens in their construction and application. This would include attention to various decision points – supplemented with very specific information about what is known from research or experience about what it takes to accomplish and sustain the goals. One important aspect is to understand this for each racial/ethnic group of interest – that is, incorporating what we know about how to improve school readiness, increase life-expectancy, support collective leadership, or increase community well-being for particular groups with particular historical and contemporary institutional, cultural and legal contexts. What is not yet known? How will we reconcile different beliefs about how change happens and is sustained based on what we know and what we don't? These questions should be negotiated collectively in order to create a more fully racialized theory of change or logic model.

It seems obvious that we would turn this lens inward to the theories of change and logic models that we use to evaluate work with racial goals. But our experience as evaluators suggests this takes real intention and some courage on everyone's part.

It seems obvious that we would turn this lens inward to the theories of change and logic models that we use to evaluate work with racial goals. But our experience as evaluators suggests this takes real intention and some courage on everyone's part. The challenges are both political and technical. For one, we need to accept that some things cannot be measured, or are not worth the expense to measure, even though they are important to do. For example, we know that very young children need to feel secure and cared for by capable adults, even if it is very expensive to capture the extent to which that occurs in a large population. That doesn't mean we shouldn't invest in strategies likely to up the number of children receiving such attention on a daily basis. We also need to be willing to separate documentation of changes from attribution of those changes to a particular strategy, organization or set of actions. In some instances, it may be enough to know that change is occurring in a positive

and important direction, and that a group is implementing its own strategies very well – so that we can make a reasonable case that they are contributing in some way to the improvements. Accepting these limitations on what evaluation can and can't do, while unsatisfying, would demonstrate some new understandings of the limits of a dominant cultural frame on our practice.

Outcomes

All of these issues also apply to defining outcomes for evaluation. Theories of change and logic models make sense as tools for evaluating work with racial goals when we can develop a set of genuinely relevant outcomes.

Because we have never achieved social or racial justice in the U.S. on any of the major indicators we care about (education, income, health, access to employment, etc.), we really do not know what it will take. Absent that experience, one way to develop meaningful outcomes would be from retrospective documentation of what worked in places or among issues where racial equity, reductions in structural racism or privilege or their consequences have occurred, particularly if those changes have endured for some period of time.

Another approach is to begin to gather the wisdom of people who can answer the question "How would you know it when you see it?" This might be a start of a common set of outcomes and indicators for evaluation of these kinds of efforts. Together, we could look for outcomes with the following properties:

- ▲ They meet certain technical considerations – for example, they are actionable (the work you are doing could affect the outcome), reliable (different people observing the same behavior would measure it as the same behavior), universal (they cover the population of interest, or the extent to which populations of interest are excluded is known) and so on.¹
- ▲ They hold a particular group or organization accountable to a reasonable contribution to a larger issue, not the whole. Or, they focus on collective contributions rather than individual ones, without necessarily assuming that a particular result must be attributed in a cause/effect fashion to any given source.
- ▲ They are easy to measure, or if not, they are worth the investment – the measuring effort might itself be part of a structural change worth pursuing (for example, tracking cognitive and emotional development of children of every racial group at ages one and three in a community; or tracking the differential impacts of employer-focused

Privilege and Racism in Evaluation

While the language of evaluation is often “race-neutral,” in many ways its methods and consequences are not.

While there are evaluators who are white and evaluators of color, the practice of evaluation itself – its fundamental assumptions about what is knowable and what is possible – draws substantially from dominant Western cultural worldviews, particularly in the U.S.

Further, evaluation is almost always applied in relationships of unequal power. Foundation staff members have power to distribute resources, but often only to the extent they can justify those decisions based on their organizations’ assumptions about what “good investments” look like. Practitioners have power to turn those investments into effective work and benefits for constituencies, but often only if they can demonstrate success against measures agreed to by others.

Evaluators can help negotiate these power differentials, but are often depending on the parties to this negotiation for some or all of their livelihood. Vastly different worldviews and power dynamics are not easily addressed in these types of relationships – including issues of privilege and racism in the work, and in the practice of evaluation itself. Unexamined assumptions about how the world works or what is important can be reflected in the evaluation timeline, as well as in what constitutes acceptable or compelling evidence of progress or success. Cultural racism and white privilege in particular can affect whose and which type of data are considered valid, or even which parties first see findings (before anyone has a chance to correct the evaluator’s errors of fact).

Generally, neither white evaluators nor evaluators of color are trained to apply a lens of privilege in evaluation to our work, though our own life experiences may bring them to the fore. And, like everyone else, we don’t know what we don’t know. Absent that, we may contribute to setting unreasonable expectations for what a group should accomplish in a given

timeframe – by, for example, failing to recognize the urgency a group feels to create a small, tangible short-term “win” before tackling a more systemic one or by discounting organizing, leadership development or other “relationship building process” goals early in a community’s work. We may also continue to assess the effectiveness of actions in the longer term using data that systematically miscounts certain racial/ethnic groups, including many kinds of census, health department, Bureau of Labor Statistics and other public data sets. Sometimes there are incentives and disincentives to accurately report information. Examples include reports on the incidence of domestic violence or nonpayment of child support where the consequences of reporting differ for people of varying legal status or racial/ethnic group.

One way to reduce unintended white privilege and racism in the practice of evaluation is for evaluators, funders, practitioners and constituents to take time together to apply a structural racism analysis to the work in which they are collectively engaged.

But even when evaluators and the other parties to evaluation believe deeper learning would be an important investment, the work to create common understandings often gets short shrift. The parties may fail to discuss or agree on reasonable timeframes for showing progress, or what types of changes in the short term are likely to predict longer-term successes, particularly in regard to multi-system structural and institutional issues. The consequences for raising these issues in unequal power relationships may make also make evaluators, practitioners or funders feel the fight isn’t worth it given other more pressing disagreements. Sometimes just the pressure to get an evaluation up and running overrides good intentions. But without these negotiations, evaluation can thoughtlessly reproduce dominant culture assumptions, make specious links among short-term observations and the likelihood of longer-term change and thus applaud effort rather than results – colluding in maintaining structural racism.

— Sally Leiderman

immigration enforcement versus worker-focused immigration enforcement in a given community).

- ▲ They are closely tied to what we know about how to achieve a particular goal, based on credible and racially explicit research or experience.
- ▲ They incorporate the most important values and defining features of the theory of change for the work being evaluated – or at a minimum, they are consistent with those values and defining features (with attention to

reconciling different theories of change among parties to the work).

- ▲ Collectively, and where possible, they represent our best understanding of the necessary and sufficient set of short-term results that predicts long-term success.
- ▲ They are strategic for a community, organization, constituent group and funder to measure (given scarce evaluation resources); accomplishing them is intrinsically important, and documenting their accomplishment will

Indicators of Weakening Structural Racism

TYPES OF OUTCOMES

Racial Equity

If progress toward racial equity were occurring, data would show that, for example:

INDICATORS (EVIDENCE): HOW WOULD WE KNOW IT WHEN WE SEE IT?

- ▲ The average life expectancy of individuals is no longer related to their racial/ethnic identity or the statistical relationship is less strong
- ▲ Rates of graduation from a four-year college or university are equal across racial/ethnic groups or moving in that direction
- ▲ The strong statistical association between family wealth and racial/ethnic identity declines

Transformative Public Policy Changes

If public policies were changing in ways that were transformative, we would see, for example:

- ▲ There would be revenue sharing among majority white and wealthy communities and majority people of color, lower-income communities in the same region for education, public safety, transit and other essential supports
- ▲ Every citizen would retain their vote, including incarcerated individuals

Transformative Changes in Narratives About Race

If people were more accurately understanding systems of structural racism, we would see, for example:

- ▲ School curricula, faith-based materials, popular media, museum exhibits, arts materials across a range of races, ethnicities and spaces reflect an understanding of the existence of white privilege, structural racism and their historical and contemporary consequences
- ▲ Everyday people speak up when public figures deny the existence of racism or privilege
- ▲ Increased system- or institution-based efforts to identify ways to reduce structural racism by that name – the idea is not taboo within public institutions

Strategies Having Their Intended Effects

If the strategies of organizations and movements were meeting their goals, we would see, for example:

- ▲ Insider strategies (changes in regulations governing actions of financial institutions, system-wide changes in standards for hiring and distributing teachers, etc.) creating some of the intended changes in policies, practices and outcomes of targeted institutions or systems
- ▲ Advocacy or other outsider strategies (economic boycotts, voter registration, public education or campaigns) garnering positive public attention, additional supporters, early “wins” of the kind intended

Collective Sufficiency of Intended Strategies

If communities or movements were putting in place everything needed to achieve their goals, we would see, for example:

- ▲ As a group, the implemented strategies work together to address system, institutional, organizational, group and individual change
- ▲ Each implemented strategy is based on research/evidence about improving outcomes for each racial/ethnic group as well as the total population
- ▲ Intensity and duration of strategies, taken as a whole, consistent with research findings on high-quality changes following best practice standards so that proven efforts are not diluted when they are replicated

TYPES OF OUTCOMES	INDICATORS (EVIDENCE): HOW WOULD WE KNOW IT WHEN WE SEE IT?
<p>Mechanisms to Anticipate Efforts to Undo Policy and Practice Changes</p> <p>If we were able to get ahead of retrenchment and resistance, we would see, for example:</p>	<ul style="list-style-type: none"> ▲ Sustained or endowed resources support tracking and sharing of outcomes at regular intervals over time. Data are collected that allow for racially disaggregated as well as aggregated results ▲ Vocal constituency can call for continued tracking and to hold leadership accountable for results ▲ Functions required to maintain these efforts are embedded in law or policy, or supported through line-item funding in appropriate budgets
<p>Collective Capacities of Organizations and Coalitions</p> <p>If what seem to be the core elements of effective movements were being implemented, those involved would together exhibit, for example:</p>	<ul style="list-style-type: none"> ▲ Mix of seasoned and emerging leadership; effective ways of supporting current and emerging leaders to strengthen their networks and capacities ▲ Each organization/coalition has an articulated analysis of structural racism and privilege, and continually shares and strengthens its analysis throughout the leadership and the base ▲ Decisions substantially driven by people of color; a large, multiracial base supports the work

motivate continued action, encourage new people to join an effort, or serve as an additional strategy for positive and transformative change.

Creating outcomes that meet these criteria sets a high, but valuable, standard.

The table above offers a range of ideas, based on various answers to the question, “How do you know that structural racism is being reduced when you see it?”, recast as potential short- or intermediate-type outcomes.² The first column describes categories of outcomes. Most are intended to signal that the work is on a path of change, before the hoped-for benefits could be observed. The second column lists some examples of characteristics or states of being that could serve as indications that those short-term outcomes are being achieved. The language is broad; we assume that people would use more specific language for an outcome based on the actual work being measured.

Conclusions and Next Steps

Evaluation practice can be improved generally, and particularly to support evaluations of efforts with racialized goals. But if we are

going to “teach to the test,” it is important to set up tests worth working towards. We do not yet know what it will take to create and maintain social or racial justice, or even racial equity of many kinds for large groups of people (at scale). It may be some time in the future before we do know. In the moment, however, we can do more, working collectively, to develop a base on which to build more useful evaluations.

One step is to rigorously review our own evaluation practices to reduce white privilege and racism in evaluation design, power dynamics, methods and consequences. Another is collectively to create a structural racism analysis of the issue being addressed with all parties (practitioners, constituents, funders and evaluators) and to use that analysis to set expectations, create outcomes and indicators and make meaning of findings.

In addition, we can all help define “how we know it when we see it” as a basis for developing more nuanced and useful outcomes and indicators to measure progress towards reduction in structural racism or its consequences. I hope that readers will consider whether they can contribute to this knowledge, by asking themselves, in hindsight:

- ▲ What “turned the corner” on an aspect of racial justice work?
- ▲ What combinations of such things seem to be the necessary and sufficient bundle of components or results that mattered?
- ▲ How did we know them when we saw them?
- ▲ What does this imply about useful short-term outcomes that predict long-term progress and achievement of the racial equity and social justice goals we strive for?

For evaluators, I hope we will bring a deeper understanding of structural racism and what reduction in structural racism would look like to theories of change, logic models, outcomes and indicators. I also hope we will bring a fully racialized lens to the practice of evaluation as it is currently constructed – helping to negotiate appropriate timelines, expectations, ways of knowing and more equitable consequences for evaluation results based on deep and collective review of the assumptions and worldviews being privileged in the work and the evaluation. For evaluators who are white, like me, this includes deeper knowledge and willingness to confront our own white privilege. Together, all of these steps might help all of us increase the effectiveness of our work and the usefulness of our evaluations.

Sally Leiderman is president of the nonprofit Center for Assessment and Policy Development. She is an experienced evaluator of efforts aimed at reducing institutional racism, supporting racial equity or building more inclusive communities, including Project Change, the Bridging Initiative of the National Capitol Region, Communities Creating Racial Equity, Communities for All Ages, the Americans for Indian Opportunity Ambassadors Program, and others. Leiderman cocreated www.racialequitytools.org and www.evaluationtoolsforracialequity.org and coauthored *Flipping the Script: White Privilege and Community Building*, including its chapter on evaluation. www.capd.org

-
- ¹ Center for Assessment and Policy Development. 2005. *What Is an Outcome and What is an Outcome Indicator? Evaluation Tools for Racial Equity*. Retrieved from www.evaluationtoolsforracialequity.org/evaluation/tip/doc/2a01.pdf.
 - ² Sources for the table include interviews with Akonadi Foundation management, staff and grantees as part of developing their evaluation framework; evaluations of Project Change and Communities Creating Racial Equity Initiatives; development work for www.evaluationtoolsforracialequity.org; www.racialequitytools.org, and *Flipping the Script: White Privilege and Community Building*. The table is also substantially informed by Maggie Potapchuk (MP Associates), Barbara Major, Sam Stephens (CAPD) and Linda Bowen (Institute for Community Peace) – partners in much of the work just noted.
 - ³ Several of these indicators were suggested by grantees of the Akonadi Foundation.
 - ⁴ One way to look at racial equity is to analyze how different groups do on an outcome of interest. As an evaluator, one could say that racial equity exists when the variable of “race/ethnicity” no longer predicts (in a statistical sense) how one fares on that outcome. That is, some people do well and some people do less well, but you can’t predict those most likely to be healthier or less healthy, or paid more or paid less, or on a board of an organization or not on the board of an organization, based on the racial/ethnic group to which they have been assigned. This definition helps to explain why racial equity is a very important goal, but not the only goal for social justice work – which may be working towards a redistribution of opportunities or power, with racial equity just one piece of that.