tcc group

strategies to achieve social impact

briefing paper

Success by Design

How R&D Activates Program Innovation and Improvement in the Nonprofit Sector

Peter York

Similar to many organizations, the Philadelphia Zoo got to a point where it needed to assess how well it was doing in meeting its ambitious goals. In the mid-1990s the zoo had crafted a new mission, which called for advancing "discovery, understanding, and stewardship of the natural world." Years went by. Exhibits were built, opened and closed, and millions of people stepped through the zoo's doors to see its carefully constructed world of animals, plants, and habitats. And yet, how far had this experience gone in shaping visitors' attitudes? Even more profoundly, could the zoo take credit for shifting people's actions, for turning them into the stewards it had envisioned?

The zoo's efforts to answer these questions suggest a new approach to learning that may no longer warrant the label "program evaluation" as it is typically used in the philanthropic and nonprofit sectors. In this new approach, the act of program measurement doesn't focus on proving something to an audience of funders who are looking for validation of an entire program's right to exist or identifying an experimentally-derived solution to complex social problems. Instead, it seeks to serve the people who create and design programs — the on-the-ground social innovators who benefit from direct insights that improve on their interventions. In other words, it functions like R&D in the private sector, providing a specific look at what is actually working and what

INSIDE

isn't, from the perspective of the target of the intervention.

This new approach has implications on both a practical and philosophical level. Most program evaluations assess impact for an entire group of people receiving the intervention, compared to a similar group that didn't receive it rather than examining who within the group benefitted, who did not, and why. A traditional evaluation would aspire to assess whether students in an afterschool program are more likely to attend college versus those in a control or comparison group who didn't participate in the program. The goal of such an evaluation is to determine if the whole program made a difference for the group that participated. If there was a statistically "significant" difference between the group averages for those who received the program versus those who did not, the program is considered a success.

Evaluations using any form of comparison group design, including pretest-posttest, do not have to examine why some afterschool program participants do not achieve the desired outcome, as long as the whole intervention group average was higher than the no intervention group. Conversely, an R&D approach assumes that every student matters and strives to understand what specifically worked for which sub-groups, with which program design elements, and with which resources. It asks how background characteristics such as gender and socioeconomic status affect the outcomes. It seeks to understand how specific program design elements may address these barriers. A traditional evaluation approach, via a comparison group design, can lead to important conclusions as to whether a program works. However, the methodologies it deploys do not force evaluators to answer the question program designers really want to address: how do you reach those who didn't succeed?

An R&D approach differs in other significant ways. Instead of looking solely at the results of an entire program, it focuses on the cause-and-effect relationships between the unique and combined program design elements and the results for beneficiaries. If a student in the afterschool program scores higher on a standardized test, was it extra tutoring or a teacher's style that made the difference, or both?

Looking at cause-and-effect starts with setting reasonable goals. It's hard to ascribe to one cause such a lofty result as a student going to college. An afterschool program might play a role, but it is arguably only one of many variables, most of which are beyond any one program's influence or control. An R&D approach looks at attainable metrics, such as whether a student attended classes more regularly or turned in homework assignments. Then it seeks to link these immediate results with specific program design elements, such as having personalized attention from a mentor.

The Philadelphia Zoo took just this sort of dramatic turn away from traditional evaluation when it engaged TCC Group to examine the various causes and effects in the zoo experience. It didn't attempt to measure, as one member of its leadership team puts it, whether a single visit led

R&D, Not Evaluation, is THE Tool for Social Innovation

R&D is about innovation. It is not about trying to create something brand new, and it is not about evaluating a finished product or service. Evaluation, as it is currently practiced, does not have the primary goal of incremental or emergent change, but rather judges the value of an intervention for a whole population. In the social sector, R&D strives to Incrementally improve the combination of elements that make up a solution in order to grow the results.

people to become conservation activists who change the world. Instead, it asked visitors questions about their attitudes, as well as their zoo experience. Were they now more inclined to save energy? Did they care more about conservation? Were these immediate results related to whether visitors had fun, or found the signs near exhibits engaging, or felt more connected to the animals after talking with zoo staff?

Altogether, the survey questions that TCC Group helped the zoo create probed no fewer than 50 different elements of a day at the zoo. And just as important, they explored who exactly responded to each one. In some cases, findings revealed that things like educational graphics or animal levels of activity made more of a difference to one subset of zoo visitors than another. TCC Group reviewed the zoo data further — looking at everything from gender to age to political leanings — in order to engage zoo leaders in a process of understanding which zoo experiences mattered most, and for which groups of zoo visitors.

"We were trying to figure out what kind of impact we were having on visitors and how they got there," says Kathy Wagner, who at the time was the zoo's vice-president for conservation and education and helped lead the effort. "We wanted to know what we achieved and what we could do better. Among the people who changed the most, what activities correlated with that change? Did they talk to staff? Did they feel an emotional connection with the animals?"

In the end, what proved most valuable to the zoo was not the data showing that, indeed, it was having an impact on people's attitudes and behavior. The greatest value came from highly detailed insights into what specifically was working and for whom — in other words, it came from the R&D process of examining cause and immediate effect. The zoo ended up with nine specific strategies that made the most profound difference for the greatest number of people.

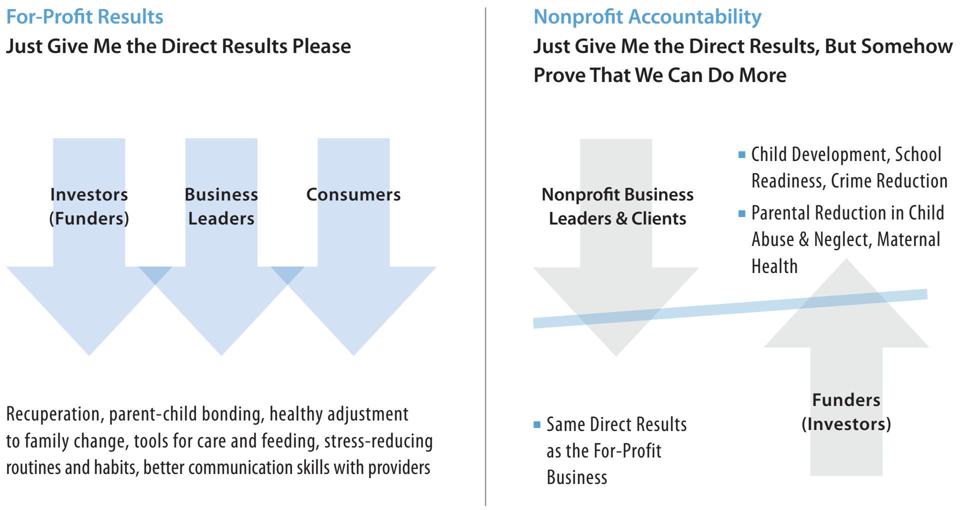
Overall, the data showed that it was critical to ensure that visitors had fun, that they learned how to help preserve natural habitats and animals, that they had numerous chances to see rare animals, and that they were left with clear connections between conservation and their daily lives. "It's not enough to talk about saving lions," notes Wagner, adding that she was surprised by how strongly "having fun" correlated with attitude and behavior shifts. "It had gone without saying," she says. "But this made us think about how we're all more receptive to messages if we're relaxed and having a good time."

The zoo was able to apply its "research" findings to the further "development" of a new exhibit, Big Cat Falls, which won a national award. "This work led us to make some wise choices about what we emphasized and what we didn't," Wagner says. For example, the data clearly showed that it was worth investing in training volunteers to talk with visitors.

The vital information that the zoo learned is akin to what a company learns from studying the reactions of people to a product and the benefit derived from it, during and after experiencing it. While R&D is not a new concept in the for-profit sector, it is new to the world of nonprofits. In

Figure 1: What Results Do Investors, Business Leaders, & Consumers Want?

Example: In-Home Support Services to New Parents, For-Profits vs. Nonprofits


For-Profit Results
Just Give Me the Direct Results Please
Investors (Funders)
Business Leaders
Consumers
Recuperation, parent-child bonding, healthy adjustment to family change, tools for care and feeding, stress-reducing routines and habits, better communication skills with providers
Nonprofit Accountability
Just Give Me the Direct Results, But Somehow Prove That We Can Do More
Nonprofit Business Leaders & Clients
Child Development, School Readiness, Crime Reduction
Parental Reduction in Child Abuse & Neglect, Maternal Health
Same Direct Results as the For-Profit Business
Funders (Investors)

fact, only five percent of the nearly 2,500 organizations that have taken TCC Group's Core Capacity Assessment Tool are engaging in R&D behaviors.

Instead, when nonprofit leaders are asked to engage in learning and measurement, they are typically encouraged, and in some cases even required, to participate in traditional evaluations of long-term outcomes. In the nonprofit sector, the "gold standard" evaluation proves whether an entire program created long-term and lasting change, versus no program at all, using some version of a comparison group study (e.g., randomized control trials, matched groups, pretest-posttest, etc.). In the private sector, R&D plays a very different role. It is a tool used by product and service designers who want to continually test new product and service enhancements to target a wider audience.

There is one major distinction between R&D and traditional evaluation. R&D is a process for improving an existing service or product to maximize the likelihood of immediate results for every individual user. Traditional evaluation attempts to prove that a service or product has changed the status of a whole group of users (i.e., not every user, but the average for the group as a whole) compared to a similar group of people who did not receive the service or product. R&D is an ongoing process, consisting of:

- Researching for whom and how a product or service development works and for whom it doesn't
- Innovating and making modifications in terms of unique and combined design elements that will help those who haven't experienced benefits
- Redesigning products or service delivery models with innovative elements and combinations of elements
- Re-testing to see if more people are directly benefitting

While traditional evaluation plays an important role in terms of gathering and teaching research-based lessons to the field, as well as garnering political and public support for investing in social change efforts, it will ultimately fall short of helping program design leaders understand and expand their success. Why? Because traditional evaluation doesn't directly help program designers examine what works by explicitly connecting the cause and effect between the groups they serve and how the program benefits its constituents using the immediate results as a way to identify and analyze why some participants benefitted while others did not.

After helping numerous nonprofits — including the Philadelphia Zoo — evaluate their programs, as well as analyzing organizational behavioral data on the learning practices of nearly 2,500 organizations, TCC Group has concluded that it is time to formalize the use of R&D, distinct from evaluation, as the program innovator's most critical learning tool.

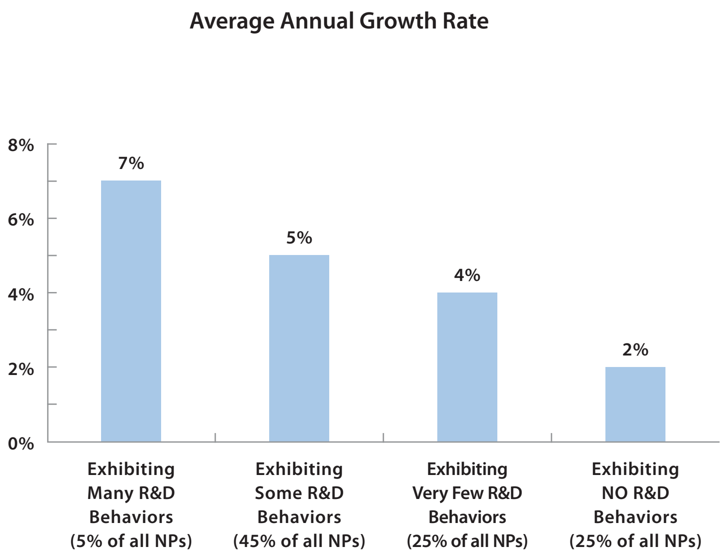
How is R&D Different?

While traditional evaluation seeks to measure return on investment, R&D actually *enables* that return by providing an understanding of how to reach the greatest number of people. How, for example, should a rural program be brought into an urban setting? How can programs originally designed for men extend to women? How can a program reach shy children as well as extroverted ones? R&D can often provide answers where traditional evaluation remains too blunt an instrument of investigation. With funding scarcer than ever, nonprofits must be able to find the answers if they are to grow. Such understanding provides a more realistic perspective from which to view costs, program design replicability, and program adaptability. And fields and subsectors can begin to develop business models that will support realistic scaling of programs.

Analysis of TCC Group's data provides statistically significant evidence that nonprofits whose leaders engage in R&D behaviors are almost two and a half times more likely to grow at or above the annual rate of inflation (refer to Figure 2) regardless of the size of an organization's budget, and controlling for all other leadership, management, adaptive, and technical capacity behaviors an organization exhibits. Specifically, the following six organizational behaviors are uniquely and significantly correlated with organizational sustainability and growth[1]:

[1] These six R&D behaviors reflect six specific Core Capacity Assessment Tool survey items that factor analyses determined measure the same construct, which TCC is labeling, "Program Design Capacity"; the Cronbach's alpha for this factor (or measurement "scale") is .78.

Figure 2: R&D Facilitates Sustainable Growth



Can R&D Improve an Organization's Sustainability, Growth, and Capacity for Program Expansion?

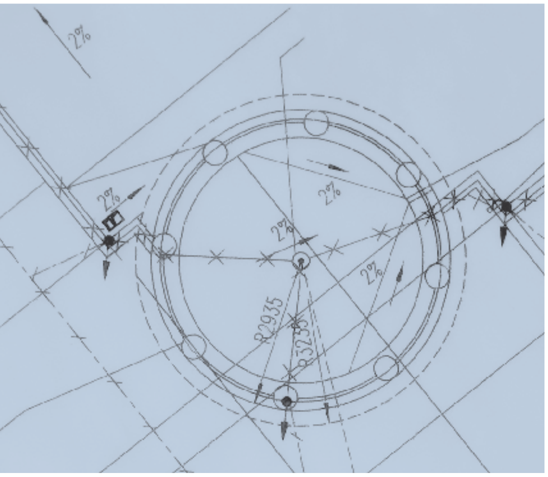
As TCC's national Core Capacity Assessment Tool (CCAT) dataset (refer to diagram, left) indicates, the answer is yes! And, the CCAT data further show that there are three reasons R&D organizations are more likely to sustain and grow: 1) R&D organizations create clear, codified, and replicable program implementation models that are delivered with consistent quality and effect because leaders learn more precisely about what works, for whom, and under what conditions; 2) R&D organizations are more effective at program management because leaders identify and develop metrics for the program implementation that differentiate the targets who achieve results from those who do not; and 3) R&D organizations are effective at identifying and cultivating new funders because they are clear about what they are funding, including being able to promise results. R&D organizations sustain, grow, and expand more effectively because they are in the leadership seat when it comes to the design of their programs. They are also less susceptible to outsiders (funders, collaborators, etc.) changing programs because leaders have proof of program success, and they manage quality more effectively.

1. Evaluating a program to figure out what works, rather than deciding if it works
2. Gathering data directly from program recipients to determine how to improve programs
3. Engaging key leaders and staff in interpreting the client-derived data
4. Determining outcome metrics by listening to, documenting, and sharing actual client success stories and results
5. Bringing design leaders together to assess and address the resources needed to deliver programs effectively
6. Leveraging R&D insights to inform the management of program implementation

R&D is not a report card, but a roadmap for those who create programs and services to constantly improve. In today's funding environment, no organization can afford not to be nimble and adaptable. Says Wagner simply, "We really need to invest more in R&D." Several key ideas distinguish its methodology and benefits, and to understand them, it is helpful to look at how the field of evaluation has evolved and what we can learn from the private sector.

Using control groups to evaluate social programs is grounded in profound shifts in the nonprofit and philanthropic landscapes. At one time, it was relatively easy to assess impact because nonprofits existed to meet basic needs; charities provided food or housing, and their impact was immediate and obvious. But as the field became more sophisticated, taking on more complex social problems, the desire grew for more advanced modes of assessing whether society was moving forward. Focus shifted away from the soup kitchen to addressing the root causes of endemic poverty. This way of thinking has produced profound insights; it is now well understood, for instance, that helping women in the developing world advance their educations and careers can lift entire families and even communities out of poverty.

But with greater complexity surrounding the design of social service programs, the task of assessing their impact became far more complex as well. At the same time, funders demanded more accountability. Increasingly, philanthropists came from the private sector and expected to see a demonstrated return on their investment (albeit a social return). For guidance on the appropriate method to determine that return, the philanthropic sector took its cues from government funders, who turned to the academic world, which relies on control group studies to assess "generalizable" impact (i.e., significant impact differences that entire populations can experience).



While a social science experimental approach used primarily for population studies can be rigorous and effective, it has also led to unintended consequences. The research goal of being able to generalize findings to entire populations resulted in metrics of success that were (and are) grand and sweeping. Nonprofits would attempt to measure their ultimate goals (or vision), which are population status changes such as "no longer homeless" or eliminating a disease. Such goals are inspirational, but they are almost always beyond the direct reach of a single intervention or program. It is more realistic, for example, to measure whether clients in a program to help the homeless actually follow through with a job referral or a doctor's appointment. Such goals may sound less inspiring but actually represent a significant achievement. More importantly, they are within reach — and they are the building blocks for achieving broader societal change, one person at a time.

Ironically, while interest in results-based accountability grew, nonprofits moved further away from a corporate business model, due in part to assuming the burden of experimental proof of big change. While Facebook might have been aiming to change the way people socialize, none of its investors demanded a study proving the attainment of such a goal; the numbers told the story as people used and directly benefitted from the Facebook experience. In the nonprofit sector, meanwhile, funders often ask nonprofits to prove their "sales pitch" rather than show how people directly benefit from a service such as job training or help managing a budget.

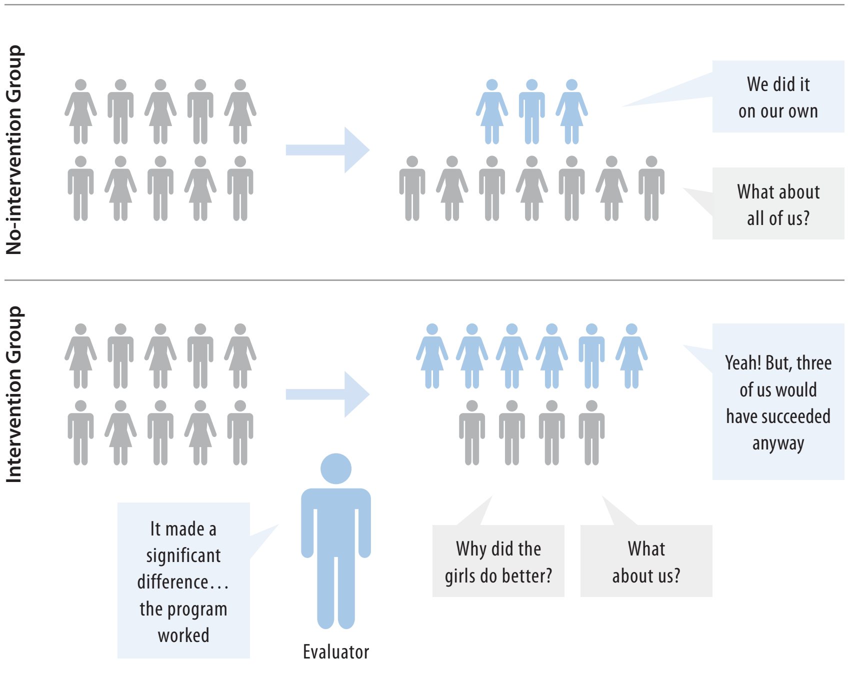
Even more problematic, traditional evaluation doesn't create opportunities for learning, which is the basis for improving programs, taking them to scale, and achieving greater impact. Control group studies are highly limited. With an afterschool program, for example, data may show that teens who participate in the program are more likely to go to college, but many questions go unanswered. Who exactly is helped? If 50 percent of participants attend college, this means a full 50 percent do not, and a control group study isn't required to say anything about what made the difference. Nor does it indicate which elements of a program or service

R&D: It's All About the Features

R&D does not suppose a product has a holistic set of unchangeable features. Instead, it assumes that there are program design elements that work for some people, in certain situations, during particular times. In the world of for-profit products and services, there are very few offerings where all of the features benefit all of the users and can therefore be sold, as is, in perpetuity. Private sector innovators constantly add, subtract, and enhance design elements to include the features that their customers want. They build in new or revised designs because they want to reach 100 percent of the potential target market.

are working. In the case of the afterschool program, is it the strong mentoring or the rigorous tutoring that improves academic performance? What works for boys versus girls? What works for kids from different backgrounds?

Figure 3: The Problem With the Comparison Group Design


No-intervention Group
We did it on our own
What about all of us?
Intervention Group
Yeah! But, three of us would have succeeded anyway
It made a significant difference… the program worked
Evaluator
Why did the girls do better?
What about us?

Here again, R&D offers an excellent alternative. The R&D approach is all about learning. A company tests a product to see who uses the product and benefits and who does not, so that it can refine design features accordingly. And it doesn't merely ask whether people like the product; it asks specific questions about what worked and what didn't. Was it this feature, that feature, or some combination? Did women and men respond differently? Was there variation across demographic lines? Were people able to recognize the product's benefit quickly enough to perceive the value of buying it? The best companies dissect large amounts of data in order to answer these questions and persuade their target audiences to purchase their products.

The Six R&D Practices

How can the nonprofit world benefit from the lessons of R&D? We have found that six main practices should guide this approach — and can help organizations reach their goals faster, for less money, and with greater engagement by their staffs and clients.

1. Focus your study on subgroups within the intervention. Comparison group studies do not aspire to understand the huge variety that exists within any group that receives an intervention. They look at the whole group's average — what percentage of the group achieved a particular outcome. What is missing from this approach is the desire to learn what works for all participants. Even if a nonprofit achieves an impressive 80 percent success rate, it is failing with 20 percent. That 20 percent is where innovation lies; it is the source of insights into how to reach an even broader number of people.

2. Investigate the cause-and-effect relationship between program design elements and their direct results on the targets of the intervention. A corollary of looking at subgroups is looking at specific program ingredients instead of the overall program. This step is critical because it enables cause-and-effect analysis that tell us what, specifically, we can do to improve the odds of success. Knowing that a whole program works is an extremely limited insight. Was it the teacher's high expectations that mattered or the extra mentoring? Only by knowing the precise causes that led to success can you refine your offering and know where to target your resources.

The Philadelphia Zoo understood this idea well. To get the most from assessing the impact of a day at the zoo, Kathy Wagner and others behind the project broke the experience down to its component parts, using R&D as a learning tool. The zoo could have stopped at simply surveying visitors about their conservation knowledge before and after their visit — and it would have found that it had achieved positive results. But it wouldn't have learned much. The experience of any program or service involves a swirl of different elements, and these can be highly nuanced. In the zoo's case, data revealed that just talking with zoo staff didn't increase visitors' conservation knowledge, but feeling a connection to their own life experiences did. Another example: seeing rare animals strongly correlated with greater motivation around conservation.

Understanding the cause-and-effect of specific elements enabled the zoo to make its exhibit, Big Cat Falls, far more effective at fulfilling its mission. Armed with new insights, zoo leadership asked, how can we build a connection between visitors' experience and conservation? One solution was a computer game that allowed visitors to help buy a radio collar for lions involved in a conservation project. "I'm so proud of what we did," says Wagner, citing how the zoo created more "keeper talks" and made them far more personal. Instead of just talking about "rhino biology," zoo staff would comment on how a particular rhino liked to have its head scratched.

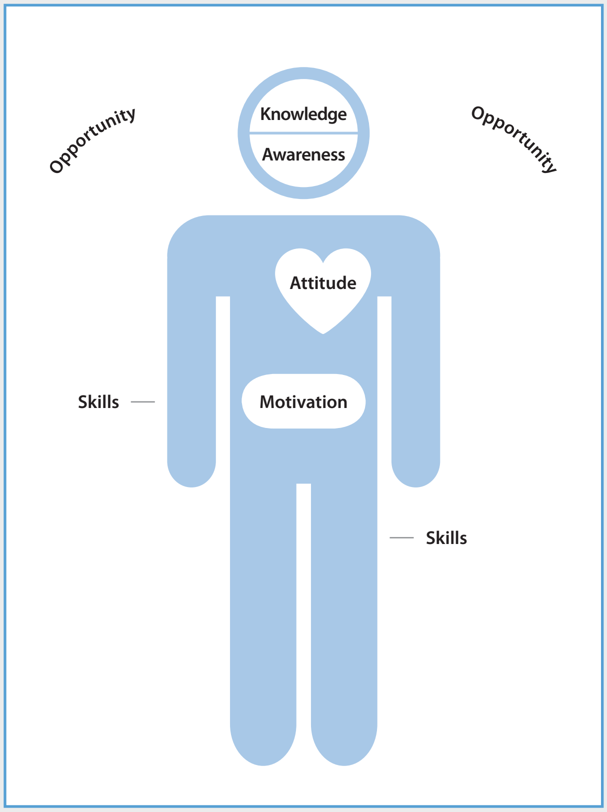
In addition, the zoo created a game for young kids in which they could become a cat, choosing, for example, whether they preferred spots to stripes. Visitors also had an opportunity to forge a personal connection by purchasing baby booties or other goods knitted by women in a village in Asia where a local organization worked to protect snow leopards (with proceeds going to the cause).

Finally, the zoo responded to the finding that staff wasn't always talking in a way that visitors could understand. Again, cause-and-effect emerged from the data, revealing that this understanding was critical to changing attitudes toward conservation. "We really worked with staff on answering questions such as 'What does this animal eat?'" recalls Wagner. "Instead of saying 'It's a highly developed carnivore with specialized dentition,' we'd encourage them to say, 'It eats meat—check out its teeth.'" The zoo could make such an improvement only because it learned that a specific aspect of the zoo experience would help achieve the effect it sought.

3. Set realistic goals and metrics for success. You can only look at cause-and-effect if the effect you're studying is attainable. If the zoo made its effect broad and sweeping — say, turning people into conservation activists — it would have been impossible to track specific causes that were under its control. And yet, even the zoo had difficulty letting go of its bold goals. In our experience, one of the most difficult mindset shifts that an organization must make to benefit from R&D is letting go of its lofty, long-term metrics of success. An organization can keep these types of goals to communicate the vision that inspires people to mobilize, but it should not use them as measures of accountability.

Wagner recalls early discussions TCC Group held with her and her team, in which we challenged the zoo's leaders to get more realistic about what a day at the zoo could achieve. Initially, they were inclined to try to measure whether people became active conservationists — organizing events and becoming change-makers in their communities. Some might ultimately end up on such a path, but it would be difficult to prove that a zoo visit could account for such wholesale transformation.

That said, realistic goals are still significant — not just in and of themselves, but because they lay the foundation for reaching larger ones. For example, the zoo surveyed people about their conservation knowledge after a day at the zoo, asking such questions as "Before your visit to the zoo today, how much did you know about the impacts of humans on animals?" It also tested personal motivation to change behavior, asking how much visitors agreed with



Figuring Out Our Direct Outcomes — It's Not As Easy As It Appears

Achieving direct outcomes for those we target is more complicated than we often realize. Even immediate behavior changes, such as motivating students to more actively participate in the classroom, don't occur without preconditions and internalized "readiness" steps along the way. It is for this reason that TCC has developed a tool to help program designers and other stakeholders understand your program's direct outcomes; we call it "AKAMSOB." Put simply, all programs aspire to achieve an ultimate behavior change for those with whom they directly intervene. But, this behavior change may not be a direct outcome your program can address. Why? Because, the path to behavioral change is complex, fraught with intervening variables, and your program may not address all the necessary steps along the way.

Let's look at a program that provides professional development training to teachers in order to increase their use of inquiry-based practices when teaching science (i.e., change their teaching behaviors) and review how the outcome chain works for achieving the "B"ehavior change using the AKAMSOB model:

1. *Science teachers need to be made **Aware** (the "A" in AKAMSOB) of what inquiry-based teaching is*
2. *Science teachers need to acquire the detailed **Knowledge** (the "K") for delivering inquiry-based instruction*
3. *Science teachers need to develop an **Attitude** (the "A") toward — or the belief in or valuing of — inquiry-based teaching practices*
4. *Science teachers need to develop the **Motivation** (the "M") to teach using inquiry-based teaching practices, which refers to understanding why the personal benefit of delivering science teaching using inquiry-based practices outweighs the personal cost*
5. *Science teachers need to develop the **Skills** (the "S"), which in most cases requires practice and feedback in order to develop confidence*
6. *Science teachers need the **Opportunity** (the "O") to deliver inquiry-based science teaching, which refers to having the resources, tools, time, space, etc. to do so*

In most cases, if any of these pieces are missing, a person will not behave differently. The AKAMSOB model can be applied to any intervention, from direct services trying to change how people act in their daily lives, to policy change efforts that attempt to influence legislation. The point is that each element of AKAMSOB needs to be assessed for those with whom you are intervening in order to: 1) determine what your program assumes exists with respect to each step in AKAMSOB; 2) determine which step is beyond the direct control of your program, given your current resources and intervention modality; and 3) determine which of the AKAMSOB steps remain and therefore serve as your program's direct outcomes or results. These remaining AKAMSOB steps need to be measured in order to engage in R&D.

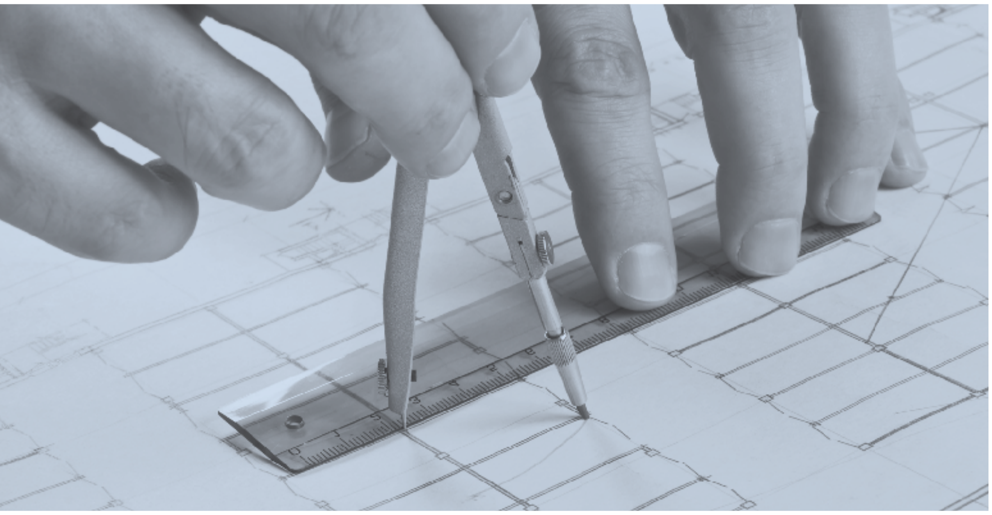
It is important to note that if you are not measuring one of the AKAMSOB steps, you are not measuring outcomes. We've been asked, "What about population status indicators such as health, economic, or educational status? Aren't these "impacts" that are beyond the control of any one program or intervention?" No. In fact, status indicators reflect the aggregate effect of individuals behaving in ways that create community-wide change. Facebook, for instance, changed the way the world socializes by virtue of the aggregate effect of the individual "friending" behaviors of each of its users.

Is R&D the Same As Performance Management?

The short answer is no. Rather, R&D leads to better performance management. This is not the same, but, in fact, cause and effect. More specifically, program design leaders use R&D to figure out what works, while program managers monitor program metrics to ensure high quality service delivery.

outcome statements like, "I will purchase products for my home and family that don't hurt the environment." While such shifts may not suggest that the zoo reached its ultimate mission overnight, they do point to concrete and essential impacts on the way people think and act.

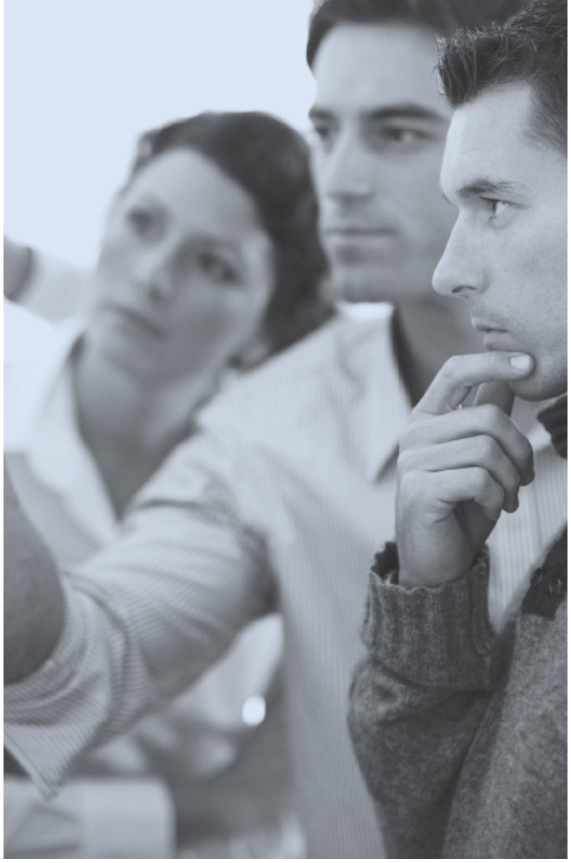
Focused on realistic goals, R&D happens more frequently at less expense, while providing real-time feedback that leads to quick improvements. A New York-based nonprofit organization also took advantage of such an opportunity and focused on training citizens to become effective policy advocates on HIV/AIDS and TB. Instead of focusing only on its ultimate goal — creating dedicated community activists who deal effectively with the media and policymakers — the organization asked TCC Group to assess simply whether its workshops were working. We designed surveys that looked at short-term outcomes such as whether the organization's use of translators to conduct sessions in French or Portuguese was effective. We quickly found that it was not — translators simply couldn't bring the sessions to life. Such a finding was not designed to assess the organization's long-term impact, but it provided the organization with valuable insight into whether its program was achieving direct results on the ground.



4. Embrace both weak and strong performance. Many organizations consider themselves successful if an evaluation demonstrates a positive impact, but to make the greatest progress, failure is equally necessary. Only by looking at both what works and what doesn't can an organization hone its strategies. Again, this idea must be considered in the context of variation within the intervention group. Whereas a comparison study draws broad findings about overall success, R&D looks for both success and failure among individuals. The fact that a particular intervention may work for white participants but not African-Americans may turn out to be the most important piece of data that a program's designers could unearth.

Jennifer Acree has applied this insight numerous times at her organization BEST in Flint, Michigan, which provides nonprofits with support to build their infrastructure. "The negative has been really valuable," she says, noting that an R&D approach revealed that executive directors participating in BEST were spending too much time in required workshops that delivered little value. BEST also discovered that while it had an impact with leadership at nonprofits, it was not reaching others at the organization, which raised significant questions about its approach and philosophy.

BEST's findings underscore a fundamental truth that traditional evaluation fails to address: no single approach will work for everyone. Social interventions, as anyone who has worked in the field knows, are as complex as the people and communities they aim to change. No two people learn — or change — in exactly the same way. Again, the private sector offers guidance on how to capitalize on this understanding. When Facebook wanted to expand its audience among the post-college market, it likely researched market trends and needs, developed web-based tools that appealed specifically to this market, and integrated the tools into the overall design of the Facebook experience. Likewise, a nonprofit like the Girl Scouts can learn through R&D which type of leadership training would work best for young women in urban versus rural areas.



Embracing the negative takes an ability to build trust, as well as confidence about pursuing a path that will best serve an organization and all of its clients. “Traditionally, you just want success points,” observes Acree, “so that you can make everyone happy and continue going forward. This model requires a lot of education with funders.” Even clients may be leery of providing feedback on what’s not working, out of fear that doing so may lead to a reduction in services. But as Acree puts it, the greatest return on investment comes from a “learning tool” that offers guidance on potential change, not just an endorsement of past successes.

5. Go straight to the source. As the Philadelphia Zoo sought to assess how well it was meeting its mission, it turned to those in the best position to offer feedback: visitors who walked through its doors and experienced its exhibits. Obvious as this may appear, many evaluations frequently devalue and even exclude feedback and reporting directly from clients. Perception-based feedback and reporting is often pejoratively tagged “self-report,” suggesting that it isn’t valid, because client-derived feedback is seen to be subjective. As a result, many in the field rely on service dosage and attendance data, even though those aren’t outcomes. Many evaluations place a premium on what are perceived to be objective indicators, which are often population-based *status indicators* (e.g., college enrollment, high school graduation, poverty, income, employment, disease rates, etc.). The evaluation metrics of success are not typically measures of immediate awareness or attitudinal, motivational, conditional, and behavioral changes. The problem with status indicators is that most are long-term social indicators that require an accumulated set of variables over a period of time to significantly change an entire population. Therefore, these allegedly objective measures are not conducive to learning for two primary reasons:

1. They don’t require directly asking clients whether they perceived receiving high-quality services (not just the required quantity). While quantity does matter as a predictor of outcomes, it is typically the case only in the extremes (i.e., you won’t get the desired outcome with the minimum intervention). The quality of an experience tends to be more predictive of outcomes.

2. It is difficult, if not impossible, to draw clear and meaningful cause-and-effect conclusions when the effect is a status indicator that is impacted by many variables beyond the control of any one intervention or program.

Even satisfaction surveys fall short of the potential learning unlocked by R&D, primarily because satisfaction is a poor measure of whether a program element or practice was specifically of high quality. Satisfaction is simply too broad, vague, and loosely interpreted. So is looking at observed behavior alone. Take kids raising their hands in class. For some, this action could signal that they’re engaged, but others might be just as engaged but simply shy. Only by getting inside the heads of participants can you get an accurate picture. Questions that probe “what has changed for you?” in terms of awareness, motivation, skill, or knowledge are far richer than simple program delivery data.

The zoo was able to draw telling conclusions based on surveys that linked a visitor’s experience to particular measures of success. It first explored, for instance, whether someone “had fun” at the zoo and then asked separate questions aimed at assessing increases in conservation knowledge, motivation to recycle, or other positive behaviors and attitudes. It could then draw connections that translated into genuine insight.

Interestingly, the zoo’s findings were based on self-perception surveys that were found statistically valid and reliable enough to pass the rigorous peer review process for acceptance in one of the field’s top academic journals, *Zoo Biology*.[2] Almost all peer-reviewed academic journals reporting on human behavior view self-perception data as valid, reliable, and relevant, as long as the survey measures meet rigorous statistical standards in terms of how the questions are constructed and interpreted.

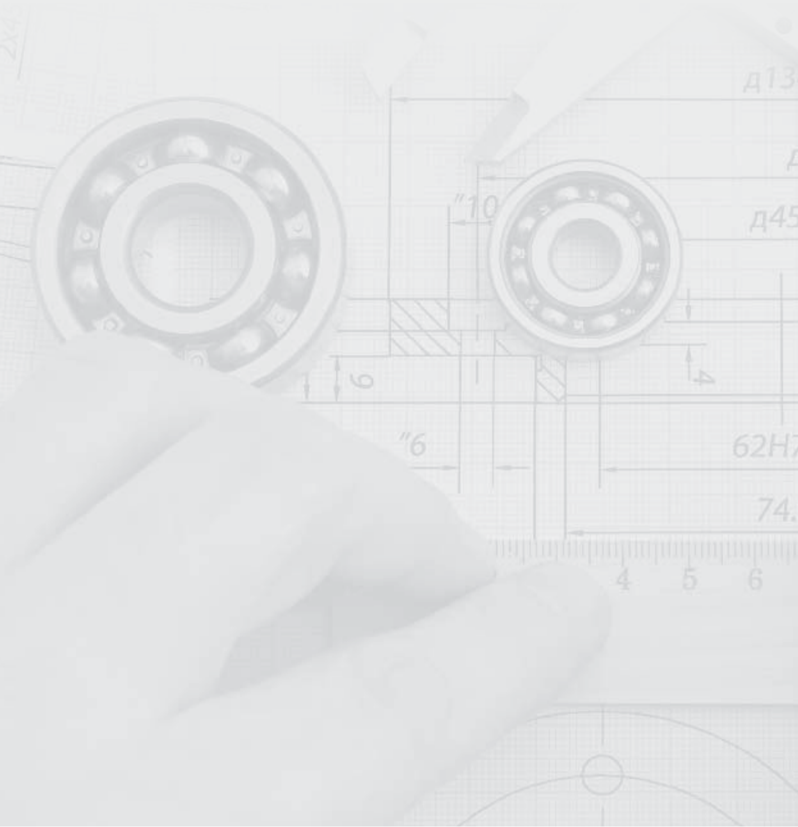
In our experience, R&D ultimately delivers more data, at lower cost, than control group studies, which tend to be more expensive to design and deliver, especially when tracking people for long periods of time following an intervention. By contrast, R&D can be effectively — and quickly — undertaken through well-designed surveys that derive richness from people’s reports of their own experience. Again, R&D calls for a mindset shift, a belief that each individual’s perceptions matter. As Wagner puts it, “If someone tells you that going to the zoo results in donating to conservation causes, you can probably believe them.”

6. View data as a beginning, not an end. As evaluators, we are all too familiar with the narrative of this type of work: collect data, draw conclusions, write and deliver a report. The dominant culture in evaluation deems that analyses and conclusions can only be “true” if derived from meticulous evaluation methodologies provided by evaluators who are “objective” with regard to the program design and implementation. R&D reverses this dynamic. Upfront stakeholder involvement in the evaluation design is not critical, except as it pertains to identifying and deciding the questions asked. The R&D process relies on technical experts in the construction of surveys and protocols, development of sampling plans, and assistance with data collection. After preliminary technical analysis of data have been completed, program leaders and designers, not the evaluator, become deeply engaged and involved in interpreting data and leading the innovation or

Is R&D the Same As Formative Evaluation?

While R&D shares many ingredients in common with formative evaluation, the explicit recipe or approach is not the same in that R&D seeks to achieve a different or improved result. More specifically, program research can only be considered R&D if it meets ALL of the following criteria:

- ✓ *Does not devote time or attention to proving that the whole program works when compared to those not receiving the program at all*
- ✓ *Always measures direct results*
- ✓ *Tries to discern which specific program design elements and/or combination of elements worked, for whom, and under what conditions*
- ✓ *Gathers quality feedback and reporting directly from the client*
- ✓ *Engages program/social innovators, designers, and interventionists as leaders for interpreting the findings*
- ✓ *Creates deliverables that are used for program design/redesign communications*



[2] Kathleen Wagner, Melissa Chessler, Peter York, and Jared Raynor, “Development and Implementation of an Evaluation Strategy for Measuring Conservation Outcomes,” *Zoo Biology 28* (2009)

re-design process. R&D is centered on an ever-changing and dynamic conversation. What's more, it puts the evaluator in a supporting role, with the real stars being those who lead the program design and implementation.

We at TCC Group believe strongly that we can provide technical support, but the team that delivers services is in the best position to evaluate what the data is trying to articulate. "A lot of the work that we do in the nonprofit sector is a balance of hard, concrete data and intuition," notes Jennifer Acree of BEST. "It's not all black and white." She adds that one of her most important tasks is poring over data to see what it's really saying, a step that she believes is undervalued by many funders. "They don't want to look at a lot of data," she says. "They want high-level, summarized information to take back to their own institutions. For a variety of reasons, they are not always on board with or able to be involved in the iterative process."

To be sure, some funders embrace the kind of learning embodied by R&D. Formative evaluation, to use the jargon of the field, accomplishes some of what we describe here. But we still see crucial differences. With formative evaluation, the evaluator usually remains the voice of authority, instead of those who truly know their target audience and understand the ins and outs of their mission and work.

At the zoo, conversations about data were highly engaging and took place in a room full of people who lived the zoo's mission every day — from the VP of education to the VP of animal programs. Having such voices at the table provided another benefit: the ability to translate what the data was say-

Does R&D Work for Policy/ Advocacy Organizations?

In many ways, R&D is the ideal approach for evaluating policy/advocacy activities, as evaluators have increasingly focused on how to assess the unique characteristics of advocacy work and its impact. At TCC Group, we have found that the high complexity and long time frames often associated with this work necessitates a focus on shorter-term outcomes. Organizational capacity, a short-term metric, is an important outcome to track—particularly for network and coalition work. Further, since many initiatives include multiple activities with different targets, evaluating the "whole program" is not an indication of progress.

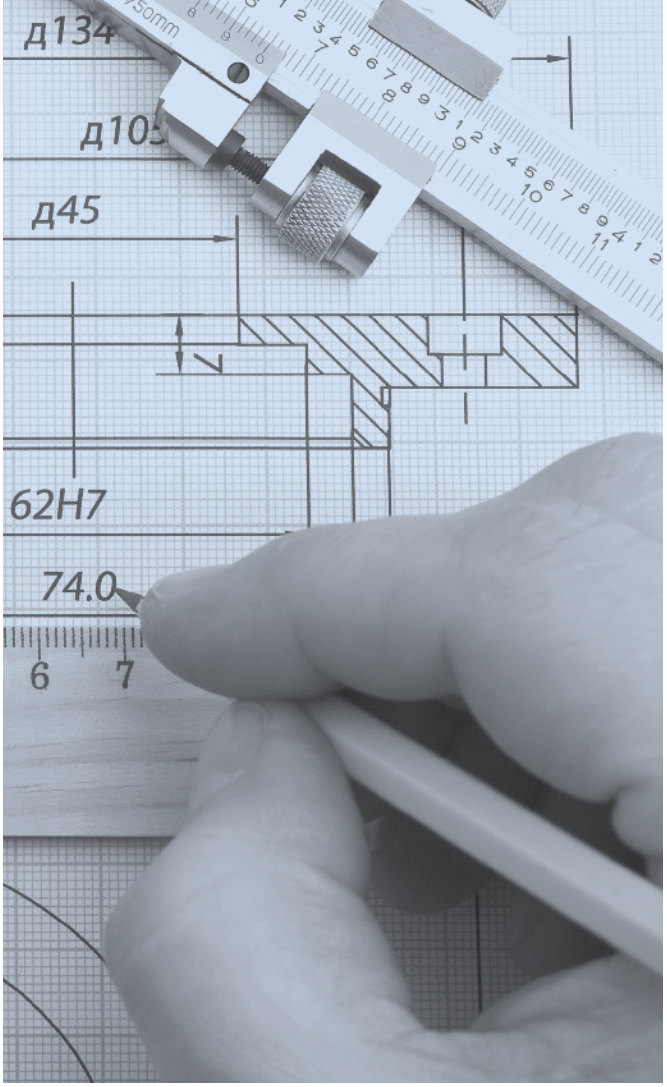
According to TCC Group Director of Evaluation, Jared Raynor, while the size of the target population in advocacy work is frequently smaller than traditional programs, the attempt to achieve behavioral change is much the same. As a result, all the general principles of R&D are applicable. Advocacy evaluation methodologies such as the Bellwether methodology, intense period debriefs, and polling data are all examples of gathering data from "program recipients" to assess progress and make appropriate changes.

Even judicial advocacy can utilize R&D. Whether through careful research on jury selection (common practice with a robust research base) or analyzing a court decision to understand whether there is room to influence through appeal or policy change, advocates can focus on measuring direct results, attempt to understand the value of various design elements, and gather quality feedback directly from the target.

ing to those who could implement changes on the ground. "We spent a lot of time trying to boil everything down to language that we could understand and agree on that we could share with our fundraisers, volunteers, and zookeepers," Wagner says.

Who Owns the Conversation?

We would go so far as to argue that having program designers "own" the conversation and the discoveries of R&D is largely what distinguishes it from traditional evaluation. Going one step further, we believe that an evaluator delivering a report cannot, by definition, take part in R&D.



The field of evaluation has been inspired by goals that can be embraced by all who work in the social sector. Who would argue with the desire to assess progress, ensure that investments are being made wisely, and ultimately, that people's lives are changing for the better? Traditional evaluations have delivered many benefits; they have challenged our thinking and pushed many organizations to new levels of achievement. While there is a critical place for this type of evaluation, it offers only limited insight for strategy leaders working on the ground who want to continually improve programs and services.

New ideas — the kind captured by the thinking and methodology of R&D — are now needed. The precise and comparative tools of evaluations that seek to assess entire populations do not lend themselves to the learning and fast program adaptations demanded by the complex even chaotic environment in which nonprofits operate. What organizations need most is the ability to collect information quickly that will translate into day-to-day actions and improvements. Indeed, these are the insights that will enable organizations to move people forward along the long, causal chain that leads to achieving the aspirational social goals tracked by many evaluations.

Learning precisely what works and for whom is the best that any one program or intervention can hope to do in the larger struggle to achieve direct and controllable results for all. By accepting this truth, and resolving to engage in learning via R&D, nonprofit leaders can gain a much more realistic and responsible means for assessing the cost of success. Once social innovators know what works and how much it costs, they can better take their programs to scale, confidently promising greater success.

Scaling is often defined as directly and uniquely causing positive, long-term social change on critical status indicators of a population's well-being — a measurement we at TCC Group believe is unrealistic and perhaps impossible. By contrast, R&D helps take social change interventions to scale by growing the proportion of those who achieve direct results, with the ultimate goal of every single participant benefitting. If growth is desirable within the context of the organization's mission and community, the R&D approach can help persuade funders and the community to support programs that have already proven successful to achieve better results.

But perhaps what is most distinctive about R&D is its intent. Traditional evaluation seeks to provide validation, and its primary audience consists of funders and community leaders who can and should hold organizations accountable. But its methodology simply doesn't meet the needs of another critical audience — the people who design, lead, and implement programs. It is R&D that speaks directly to this group and helps them reach a different goal: not that of assessing services after the fact, but of working on continually improving the actual delivery and direct benefits of those services to the people who need them.

Our Evaluation Work

At TCC Group, measurement for learning defines our philosophy and approach to evaluation — learning that enhances an organization's capacity and effectiveness, improves programs, and leads to greater impact among those in need. The real value of measurement, we believe, comes from being deliberate in its purpose and intended use.

Our Core Services

An R&D Approach — Our process ensures that organizational leaders and key stakeholders receive and use accurate, timely, and rigorously collected data to not only determine "if something worked," but most importantly, pinpoint the specifics of "what worked, for whom, and under what conditions."

Learning Systems Design — We assist in designing learning systems, frameworks, and processes to address an organization's information needs. Taking an inclusive approach, we assess a nonprofit's capacity to implement data collection analysis and maximize design leaders' ability to achieve results. We provide solutions to improve and enhance access to — and use of — data.

Cluster Evaluation — TCC Group has experience developing and conducting cluster evaluations of multiple programs across many organizations, usually as a part of a major funding effort. Typically, cluster evaluations derive "big picture" findings to inform an initiative's ongoing design and planning.

Evaluation of Capacity-Building Initiatives — TCC Group is a leader in the field of evaluation, with specific experience measuring major capacity-building initiatives for the nonprofit sector. TCC uses a widely tested, specialized set of methods, including the online Core Capacity Assessment Tool (CCAT), to measure organizational capacity-building strategies as well as community capacity to achieve greater impact.

Evaluation of Policy / Advocacy Initiatives — TCC Group is among a select group of evaluators engaging in the assessment of policy and advocacy efforts. While measuring the results of such initiatives may seem abstract or complex, TCC has created effective models to document progress and results in both domestic and international settings.

Evaluation of Collaboratives — TCC Group has developed an innovative approach to evaluating networks, coalitions, and collaboratives that takes into account the necessity of measuring not just outcomes, but also partnership performance and engagement.

Contact TCC Group

New York
31 West 27th Street
4th floor
New York, NY 10001
212.949.0990

Philadelphia
One Penn Center
Suite 410
Philadelphia, PA 19103
215.568.0399

San Francisco
225 Bush Street
Suite 1600
San Francisco, CA 94104
415.439.8368

Website
http://www.tccgrp.com

Email
info@tccgrp.com

About TCC Group

For more than 30 years, TCC Group has provided strategic planning, program and grants management, evaluation, and capacity-building services to foundations, nonprofit organizations, corporate community involvement programs, and government agencies. In this time, the firm has developed substantive knowledge and expertise in fields as diverse as education, arts and culture, community and economic development, human services, health care, the environment, and children and family issues. From offices in New York, Philadelphia, and San Francisco, the firm works with clients nationally and across the globe. Services include business planning, organizational assessment and development, research, feasibility studies, organizational evaluation, board development, restructuring and repositioning, as well as grant program design, measurement, and management. TCC Group has extensive experience working with funders to plan, design, manage, and evaluate initiatives to strengthen the capacity of nonprofit organizations.

Acknowledgements

Peter York, Senior Partner and Director of Research, would like to thank the following clients and colleagues for their valuable contributors to this paper.

- **Dr. Andrew Baker,** Chief Operating Officer, Philadelphia Zoo
- **Kathy Wagner,** formerly of the Philadelphia Zoo
- **Jennifer Acree,** Project Director, The BEST Project
- **Richard Mittenthal**
- **Paul Connolly**
- **Chantell Johnson**
- **Jared Raynor**
- **PeiYao Chen**
- **Michele Garvey**
- **Andrew Ahlquist**

Published July 2011