

# Taking the Evaluation Leap: Lessons from Urban Alliance's Six-Year Randomized Controlled Trial

by Eshauna Smith

Completing a randomized control trial (RCT) is an arduous process. For Urban Alliance, the six-year journey proved rocky and came with a cost. But, as the author explains, there is no better evaluation for answering the question on every philanthropist's lips: "How do you know your program works?"

IT'S A FAMILIAR CONVERSATION. WHEN ASKED how they know their model works, a nonprofit or foundation will tell you a story. They'll say that seventeen-year-old Darren turned his life around after going through their program, and eighteen-year-old Kaitlin discovered a hidden reserve of strength. They'll tell you about the shy girl who came out of her shell, and the unmotivated boy who found a new direction in life.

Such transformative tales are a vital tool for illustrating a nonprofit's value and impact. Here at Urban Alliance—a Washington, DC-based nonprofit that partners with businesses to provide high school students at risk of becoming disengaged from successful career or college pathways with internships, mentoring, and job skills training—we love to tell our interns' stories. Stories like that of Baltimore teen Shaquille, who struggled to support himself during high school while trying to plan for his future. We found him a paid internship with Legg

Mason, helped him to graduate from high school, and trained him in the skills needed to not only continue working at Legg Mason as a full-time employee after graduation but also go on to college.

But we have another powerful tool in our belt. Unlike most nonprofits, we can say that we have rigorously challenged, objective proof that our model works. This summer, we completed a \$1.2 million independent, six-year randomized controlled trial (RCT). According to the Social Impact Exchange, only 2 percent of nonprofits have completed an RCT, often referred to as the gold standard of program evaluation.<sup>1</sup> Whether it's the upfront cost of mounting such a rigorous study, the hidden costs to staff and stakeholders, or the potential cost of going through the process without any results to show for it, nonprofits are understandably hesitant to commit to an RCT.

We were fortunate enough to come out of the process with positive results.

Our RCT found that going through Urban Alliance's flagship high school internship program has a statistically significant impact on the likelihood of young men attending college,<sup>2</sup> the likelihood of male and female midlevel students (2.0–3.0 GPA) attending a four-year college, and the likelihood of comfort with and retention of critical professional skills over time. We now have a clear picture of what we're doing well and what we need to improve upon—and we have an empirical argument to take to job partners and funders that our model works and should be scaled out to reach even more students.

We've always known from internal data and the students we work with that we're doing something right, but completing an RCT has given us a persuasive new piece of evidence to share with those outside the world of youth-focused nonprofits—where facts often outweigh passion, and numbers outweigh anecdotes.

From 2011, when we began this study, to now, we've broadened our base from Washington, DC, and Baltimore to incorporate Northern Virginia, as well. With the addition of a new presence in the Midwest/Great Lakes region, based in the Chicago metropolitan area, we've expanded our imprint to become a national organization (not to be confused with National Urban Alliance, a different organization). The study's interim report, released in 2016, was also leveraged to win an Investing in Innovation (i3) validation grant from the Department of Education—one of just fourteen awarded in 2015—which will help our effort to expand to a fifth location in fall 2018. Our internal evaluation has also become more sophisticated, increasing from one full-time staff member dedicated to internal evaluation work to three, with work already begun on a second RCT to study our program's impact across all four current locations.

Our results justified the arduous RCT process, because we'll be using what we've learned in order to improve—and, most important, expand—our program, ultimately allowing us to serve more students. But the process was by no means smooth or without cost, and we learned a lot along the way. We didn't avoid the trap—which nonprofits often fall into with an RCT—of jumping in without first honestly assessing our readiness for such a venture. So for nonprofits thinking about completing an RCT in the future, we want to share what the process really entails and offer up some hard-won advice.

### What We Did

Over the past decade, the philanthropic sector—from government agencies to foundations to nonprofits—has been asked the same daunting question: How do you know it works? On the surface, the question makes sense. Resources

are limited, so investments need to be strategic. Let's build out the interventions that work and change the ones that don't. But for better or worse, this proof point has evolved. Collecting your own data is necessary, but insufficient. Stakeholder surveys and internal assessments may signal a more sophisticated nonprofit evaluation system, but they don't answer all questions. External evaluations, particularly ones that are designed to get at issues of causality through impact experiments, are now all but required.

In 2010, Urban Alliance received funding from Venture Philanthropy Partners, through the coveted Social Innovation Fund (SIF) from the Corporation for National and Community Service, to help us bolster and expand our program. But eligibility to receive funding required a third-party evaluation. Our independent analysis was conducted by the Urban Institute, and the full evaluation process consisted of two parts: first, a process evaluation, in which the researchers examined the program's delivery via interviews and observation; and second, an impact evaluation (in our case, an RCT) to measure how much bearing our program had on our students' success. Outcomes of students who had been offered access to the program were compared to those of a group of similar students who had not been offered access, by controlling for unobservable factors (such as student motivation) that could impact results. The Urban Institute used a randomized lottery to assign applicants to either the treatment group (the group with access to the program) or the control group. It was cold, but fair.

The Urban Institute followed the students and control groups in the 2011–2012 and 2012–2013 classes, measuring the program's impact on college enrollment and persistence, comfort with hard and soft skills, and employment and earnings, among other factors.

## Challenges

**1. External relationships.** Urban Alliance has always prided itself on the strength of its partnerships. Over the years, trust, open dialogue, and a mutual passion for helping underserved students has created a strong relationship between our staff and the counselors and principals of Southeast DC and Baltimore. But when in the fall of 2011 we began recruiting not just to fill our 2011–2012 class but also to fill the study's control group, we were essentially recruiting students we knew we wouldn't be able to serve. Given that our aim is to give opportunities and an expanded sense of possibility to youth from underserved communities, from the outside it appeared counterintuitive and even cruel to reject the very students Urban Alliance was created to serve for the purposes of this evaluation. Our long-term objective—to use (hopefully) positive results to serve more young people overall—was obscured by the short-term disappointment we caused students.

We mistakenly assumed that our partners would see the potential of this research just as clearly as we did. Families and counselors were understandably upset—but we hadn't foreseen that consequence. For a partnership organization like ours, a study's success relies on one large but little-discussed caveat: that partnering schools and districts will want to participate. These partnerships worked so well in the past because they were mutually beneficial. We needed their students to run our program; they needed our program to support their mission. The RCT changed the terms of that partnership, because we could no longer guarantee spots for students identified by their counselors as most in need of intervention. As a result,

# NPQ

“NPQ is a courageous journal in a field that will need courage.”

—Jack Shakely, NPQ reader

Thank you for subscribing to NPQ!

We see ourselves as being in deep partnership with you, our readers.

We rely on your feedback, your survey responses, your stories for our editorial content. Subscribers are the lifeblood of our organization *but we also rely on your donations for our financial health.* We keep the cost of

our subscriptions low—

we don't want cost to be a barrier for anyone! But if you can give more—and if you value what NPQ has provided for more than fifteen years—consider joining a growing group of your fellow readers, and go to [www.nonprofitquarterly.org](http://www.nonprofitquarterly.org) to make a donation today.

—Ruth McCambridge,  
Editor in Chief

some schools told us to come back next year—after the lottery. Some told us not to come back at all. We were accused of chasing money, or of sacrificing our values. We were asked how we could still claim to be pro-student if we were rejecting some of those who needed our help the most. It was a fair question—and one that we were not ready to answer. We would also not have empirical evidence to support our answer—when we had one—for another six years.

2. **Our team.** The challenge of recruiting students from uncertain partners placed an added burden on our staff out in the field. Evaluators obsess over a study's sample size. The larger the sample, the easier it is to attribute impacts to the program's intervention and not just to chance. But we found that it was much simpler to plan for a large sample size than to actually reach it. Recruiting students for the study meant doubling the normal effort required to fill one of our classes but in the same space of time our program staff were used to having. And the skepticism we faced from school partners about undertaking this evaluation only made the task more difficult.

Thus, the RCT also measured the psychological impact of these challenges on our staff. Though most of us had in mind the long-term benefits of the study, our program staff were on the front lines of the process and interacting with disappointed stakeholders every day. Furthermore, most people who choose to work in youth development do so to give—not deny—assistance to young people in need. Many staffers were disappointed. Some became disengaged. Some even left. As an organization, we anticipated certain growing pains, but the internal impact of this type

of large-scale evaluation was unexpected. The greater good argument will always be controversial, but we underestimated just how much of a strain it would put on team morale.

3. **Feel-good stories versus real numbers.** It's easy to feel good about your work when you see the individual stories of achievement among your clients. But upbeat stories are very different from cold, hard data.

We were fortunate to see statistically significant results, as many nonprofits go through the RCT process only to get null or even negative results. But after all the negotiations, concessions, and heartache, not getting empirical confirmation that the results we saw on an individual basis translated into massive numbers was disheartening. Agreeing to have external evaluators look under the hood is one thing; challenging decades' worth of core beliefs is something else. Our inexperience with this type of evaluation led us to overlook the possibility that our results wouldn't confirm all our biases. But the more unexpected results—for example, a positive impact on young men attending college but not on young women—mean that after twenty years of perfecting our model, we still have a lot of room for improvement. And that's as it should be. These specific results will now help to guide us as we grow as an organization, and ultimately will help to make us more effective down the line.

## What We've Learned

We came out of the RCT process relatively unscathed. We have positive results to show our partners and a powerful argument to make for expanding our program. But there's a lot we wish we had known from the outset. Before undertaking a large-scale evaluation like

an RCT, a nonprofit should be prepared to do the following:

1. **Staff accordingly.** Implementing an RCT is a process with many moving parts—from doubling recruiting efforts to managing relationships to keeping staff motivated and informed, and so much more. And an internal staff member needs to be at the helm throughout the study period to ensure that everything is running smoothly and no ball is being dropped.

A stand-alone evaluation staffer is a luxury for most growing nonprofits, especially those with still-nascent performance and accountability systems. That evaluation function is usually shared across departments, with, for example, the development team collecting statistics and the program officers handling demographics. But without someone fully devoted to the task, a large-scale evaluation can easily go awry.

For example, an external evaluation requires a mountain of paperwork: parental consent forms, memoranda of understanding, institutional review board forms, and so forth. Some paperwork is to be expected when working with underage students and job partners, but the amount grows exponentially when you factor in an RCT. And the timeline for completing all this extra documentation is often truncated, since paperwork must be in place before the study can commence. Any delays could put the entire evaluation in jeopardy. When one school district pushed back their institutional review board (IRB) decision, Urban Alliance almost lost an entire year of observations. It took relentless phone calls, wrangling, and sheer stubbornness to get all the final documents signed.

An internal evaluator plays another key role: liaising between external

evaluators and program staff. Countless decisions must be made during the evaluation process, from how to interview the staff to how best to observe program delivery and conduct client focus groups. To keep the program running smoothly during this time, these components should be conducted as unobtrusively as possible. A staff evaluator's inside knowledge of how the organization works makes this process much easier.

A dedicated evaluation staffer also brings the subject-matter expertise necessary to push back on methodological decisions made by the external evaluators. Decisions on statistical power, approval of survey questions, and agreeing on outcomes of interest are critical to a study's success. It's best to have someone on staff with an understanding of how such evaluations work to help make these choices.

2. **Establish a strong internal performance measurement system.** Impact evaluations and performance measurement are used to answer different questions, so one doesn't completely replace the other. Performance measurement tells us what our intervention is doing; impact evaluations like an RCT try to demonstrate what is happening *because of* our intervention.

Strong performance measurement activities are ongoing and can be completed much more quickly than impact evaluations, which can take years. This quicker turnaround allows for real-time course correction, while a longer-term study informs the program with respect to the bigger picture. Despite these differences, performance measurement should be considered a prerequisite for an RCT. This smaller-scale measurement will identify gaps in implementation and

delivery and underperformance of both staff and client outcomes. And, if they're present in the internal evaluation, they'll certainly appear in the external evaluation.

Nonprofits can use such internal performance measurement to work out any kinks in their model before inviting deeper scrutiny. Implementing a robust performance measurement system also helps to test whether the outcomes the nonprofit wants to see in an RCT are even attainable or observable. It's reasonable to challenge an external evaluation design that hopes to test the intervention's impact on a certain outcome if the nonprofit knows it will be impossible or even inconsistent to collect. By taking the time to experiment, nonprofits can get ahead of potentially null or negative results.

3. **Overcommunicate internally and externally.** Too often, nonprofits get caught up in the excitement of winning substantial funding and overlook the smaller details of executing a grant's required external evaluation. The first thing that usually gets lost in the shuffle is informing stakeholders.

A rigorous study will necessitate significant procedural changes, not just for the nonprofit but also for external partners. The onus is on the nonprofit to fully explain these changes and how the students and partners will be impacted, and set a timeline for how long these changes will be in effect. But all that preparation can be overwhelming without a clear and compelling explanation of the study's benefits.

As illustrated earlier, we did not recognize the chilling effect the randomized lottery would have on our partners. Making it clear that a disruption is temporary and controlled can soften the news. And helping

your partners to see the value not only for you but also for them will help to ease strained relationships. Communicating clearly to your partners and other stakeholders what positive RCT results will mean for what you can do for their clients and communities in the future will help to mitigate some of the frustration up front. The front-line program staff need to be well versed in these talking points from the get-go. Discrepancies in your internal messaging—including the value that consistent messaging brings—will echo externally.

Additionally, clear communication can only go so far without the right tone. Youth development is a field grounded in empathy, and that can't be forgotten when communicating what will be disappointing news to

many partners. Understanding their frustrations while presenting the silver lining is essential to making sure partners feel heard and valued during what is always going to be a difficult time.

4. **Fully commit.** The decision to undertake an RCT should not be made lightly. As you can tell from Urban Alliance's experience, there are costs as well as rewards to this kind of evaluation. Debate needs to be had, and input needs to be heard. But once you commit, you need to commit completely.

A full commitment requires giving in to the process for better or worse. It is not pleasant to have someone on the outside auditing your organization, but any feedback, whether positive or critical, should be welcomed as a learning opportunity. Most nonprofits are never scrutinized this closely, so an RCT is an invaluable learning tool for the groups that choose to use it. Too much time and money have been invested—and too many relationships have been tested—to ignore the results at the end of such a grueling process.



At Urban Alliance, we've certainly been tested by the RCT process. However, we came out with positive results and a compelling data set to support further expansion and enable us to serve more students. The ups and downs were ultimately worth it, because we can now increase our reach to provide critical work experiences and support to young people who might not otherwise have such an opportunity. And, as part of the i3 grant our initial set of RCT results made possible, we're now in year two of a second RCT to evaluate our impact across our current four locations.

Ultimately, if your organizational mission can significantly benefit from an RCT's external evaluation, then consider taking the leap—but make sure you're prepared for the roller coaster ride it will inevitably become.

#### NOTES

1. *Executive Summary on the State of Scaling Among Nonprofits* (New York: Veris Consulting and the Social Impact Exchange, 2013), 4.

2. See *Pathways after High School: Evaluation of the Urban Alliance High School Internship Program* (Washington, DC: Urban Institute, 2017), xxiii–iv, [theurbanalliance.org/wp-content/uploads/2017/08/08292017.pdf](http://theurbanalliance.org/wp-content/uploads/2017/08/08292017.pdf). As the report explains, “In general, females were more likely to graduate high school than males and more likely to attend college. We found that the program had no impact on college attendance or persistence for females, but it had strong impacts for males. On each of these measures of college attendance or persistence, males in the program showed outcomes similar to females, indicating the program helps close the educational gap between females and males. For example, approximately 70 percent of males in the Urban Alliance program attended college, similar to females in either the program or control groups, but only 55 percent of control group males attended college.”

**ESHAUNA SMITH** is CEO of Urban Alliance, an organization that partners with businesses to empower underresourced youth to aspire, work, and succeed through paid internships, formal training, and mentorship.

To comment on this article, write to us at [feedback@npqmag.org](mailto:feedback@npqmag.org). Order reprints from <http://store.nonprofitquarterly.org>, using code 240308.

**THE Nonprofit QUARTERLY**

The latest news and analysis about the nonprofit sector from the *Nonprofit Newswire*

■

Regular feature articles

■

Subscription information for the print magazine

■

For more information from the *Nonprofit Quarterly* go to [www.nonprofitquarterly.org](http://www.nonprofitquarterly.org)