

Building Equitable Evidence of Social Impact

Nonprofits and funders serious about building equity must ensure that constituents are true participants in evaluating impact. Our research found that a holistic approach, blending participatory methods with empirical metrics, not only surfaced causal links to outcomes but also influenced more-inclusive organizations and culture.

By Lymari Benitez, Yessica Cancel, Mary Marx, and Katie Smith Milway



Pace Center for Girls provides girls and young women an opportunity for a better future through education, counseling, training and advocacy. Founded in 1985, Pace provides free year-round middle and high school academics, case management, counseling, and life skills development in a safe and supportive environment that recognizes girls' past trauma and builds upon their individual strengths. Dedicated to meeting the social, emotional, and education needs of girls, Pace has a successful and proven program model that has changed the life trajectory of more than 40,000 girls and is recognized as one of the nation's leading advocates for vulnerable girls.

MilwayPLUS advises philanthropy and nonprofits with a focus on research, publishing and content strategies for funders, and innovation and growth strategies for content creators. Katie Smith Milway, principal of MilwayPLUS and a senior advisor at **The Bridgespan Group**, has a background in journalism, nonprofit management, strategy consulting, and governance. Her work in sustainable development across four continents includes program design and measurement, donor development and communications, and organizational effectiveness.

ABOUT THE RESEARCH

From February to March 2021, Pace Center for Girls and MilwayPLUS surveyed and conducted focus groups with three cohorts of nonprofits that embraced participatory measures. Pace and Milway co-curated focus-group participants with Fund for Shared Insight, Feedback Labs, and Project Evident. In addition, the research team reviewed the evaluation literature and interviewed evaluation experts at other nonprofits, funders, and measurement intermediaries.

The research team's objective was to test a hypothesis, based on Pace's experience, that participatory measurement could link to life outcomes and, in addition, generate organizational and cultural benefits. Research findings supported the hypothesis, while underscoring the importance of blending participatory methods with empirical measures to build more-equitable, robust evidence of program impact.

Focus group participants were Accountability Lab (Mali), Boston Uncornered (MA), Boys and Girls Clubs of the Peninsula (CA), Gift of the Givers Foundation (South Africa), Lake County Build a Generation (CO), My Life My Choice (MA), Nurse-Family Partnership (CO), Our House (AR), Pace Center for Girls (FL), SOLE Colombia (Colombia), Think of Us (DC), and Union Capital Boston (MA).

Survey participants were, in addition to all focus-group members, Youth Truth (CA), Center for Employment Opportunities (NY), and Doorways NWFL (FL). The team conducted interviews with data-driven nonprofits Youth Truth, Center for Employment Opportunities, and Harlem Children's Zone, and with evaluation experts at RTI, Blue Meridian Partners, CUNY School of Professional Studies, and Ford Foundation. All quotations in our report derive from original interviews or focus-group comments.

Pace Center for Girls, a Florida-based multiservice nonprofit serving middle- and high-school-age girls with histories of trauma, faced an ethical dilemma several years ago: the organization and the community it serves, as well as funders and policy makers, sought concrete evidence that Pace was getting results. So the nonprofit launched a **randomized controlled trial** (RCT) to assess whether performance was better for girls in the program than for those not enrolled. But that meant withholding services from some girls (the control group) and referring them elsewhere, at odds with Pace’s mission. The longitudinal RCT would also be costly and labor-intensive—and take years—while approaches at Pace and in the field naturally evolved.

Despite the downsides, Pace pursued the RCT. Conducted from 2012 to 2018, it revealed that Pace girls were nearly twice as likely to be on track to graduate from high school as girls not at Pace. But the findings had limited application. Because the RCT was designed to evaluate progress against standard measures, like attendance and grades, it could not establish a causal link with Pace’s signature individualized services—such as counseling, anger and stress management, and building self-efficacy and self-advocacy—that played a potentially outsized role in setting girls up for success in life.

Flash-forward three years, and Pace has put into practice empirical research that can be done holistically and equitably. Instead of relying on a multiyear RCT for ultimate answers, Pace blends empirical findings with participatory methods to learn how processes, policies, and social institutions help or hinder the girls it serves. By collecting actionable information from constituents, Pace’s approach to evidence building has evolved from conducting one arm’s-length study at a time, which provides data in hindsight, to sustaining an ongoing, integrated process that directly involves girls, community members, and other stakeholders in designing research questions and analyzing the data they produce.



Pace Center for Girls’ participant-centered approach to creating evidence of impact empowers Pace girls to influence the juvenile justice system itself. Pace girls seen here in a mock bill session in Tallahassee, Florida, preparing to testify to state legislators. (Photo credit: Pace Center for Girls)

In building its research and evaluation department, Pace has uncovered new patterns of the girls' experiences by segmenting their responses according to race, age, length of stay in the program, and other variables. In the process they have identified causal links between feedback on instructor relationships and educational outcomes for the girls. Pace now immediately incorporates participants' insights into program improvements, thereby reinforcing in real time the nonprofit's culture of responsiveness and its participants' self-efficacy and self-advocacy.

Pace is not alone. It recently worked with **MilwayPLUS** and three ecosystem leaders in participatory measurement—**Fund for Shared Insight**, a funder collaborative that, since 2014, has used **grants alongside other strategies to accelerate the use of feedback**; **Feedback Labs**, which builds NGO peer-learning networks; and **Project Evident**, an adviser on evidence strategy—to conduct focus groups, interviews, and a survey with 15 organizations (including Pace) that emphasize participatory approaches in program design and measurement. Research findings fueled a hypothesis, based on Pace's experience, that participatory measurement could link causally to outcomes and, in addition, generate organizational and cultural benefits. (See "**About the Research, page 2.**") These nonprofits span a variety of fields, and most have built participant feedback into their measurement systems within the past six years, several with the help of grants from Shared Insight and peer learning with Feedback Labs. They consistently reported that the shift had increased their momentum and inclusiveness in decision-making; connected measurement to outcomes that matter to program participants; and empowered participants to influence unjust systems to move toward greater equity. None of this work obviated the need for experimental tests, including RCTs, to prove specific effects (such as **how well an activity boosted participants' health**). But the organizations did find that participant-informed measures bolstered their ongoing insight and adaptability, which was tested through the tangle of health, racial, and economic crises in 2020.

Foundations are also making the shift. **The William T. Grant Foundation** advocates using a lens of critical race theory, which calls out methods that can mask harm and dehumanize people in the name of objectivity.¹ The **Ford Foundation**, committed to addressing inequalities, doesn't start with RCTs and is continuously thinking about how to involve participants throughout the evaluation process. Even research giant **RTI International** is moving to blended measures. "I don't consider an RCT [on its own] the gold standard now," says Dr. Jennifer Laird, RTI senior manager for education research. "A more participatory evaluation is really important, and you can merge the two."

“[Focus group members] reported that the shift to participatory measures had increased their momentum and inclusiveness in decision-making; connected measurement to outcomes that matter to program participants; and empowered participants to influence unjust systems to move toward greater equity.”

Using Participatory Measures to Understand Outcomes and Impact

A core difference between RCTs and participatory approaches is at the heart of the equity argument: RCTs follow a treatment and control group over time—and then look back, running regressions, to analyze change. Because RCTs have roots in scientific experimentation, they are not designed to adapt the intervention in response to participant feedback, as the participants are subjects in the study. Nor are they designed to identify the systems and structures that create disparities in the first place.

In contrast, participatory tools—feedback, surveys, focus groups, testimonials, diaries, participant councils—derive from a social science method called **participatory action research** (PAR), which connects immediate learning with continuous improvements to programs and policies, with participants seen as experts in their own experience.² This doesn't render such methods immune to implementation in ways that could reinforce a dominant view. But PAR, coupled with a mindset for advancing equity, can empower participants to challenge inequality and foster field-level action and inclusive policies.³ PAR's cousin, **participatory evaluation** (PE), which gives program participants, staff, and other stakeholders ownership in designing and managing the evaluation process itself, emerged in the late 20th century as a subfield of program measurement, particularly outside the U.S. and among international relief and development nonprofits. PE radically shifted how to gauge social programs' effects on participants and their sense of power, always asking, "What answers are we seeking? Why? By whom? For whom?"⁴

Today, practitioners and their funders have increased the use of PE and other equity-serving and user-focused methods to account for specific contexts, such as feminist or indigenous. However, some entities, especially government funders, hold experimental methods in higher regard and consider them essential to making decisions to fund program expansion. The work to reposition the use of experimental methods, like RCTs, and expand and integrate the use of participatory approaches to create truly holistic measurement raises several key questions:

- 1. Who needs the proof?** Large-scale funders often suggest multiyear experimental studies, but the nonprofits we talked to put their participants' views first. Mark Culliton, cofounder and CEO of **Boston Uncornered**, a nonprofit that stipends core influencers in Boston gangs to train as change agents, described a correlation between reduced violence in one neighborhood with the presence of trained core influencers: "It's not stuff that a funder would necessarily focus on, but we're much more interested in what the core influencers believe." Kwame Owusu-Kesse, CEO of **Harlem Children's Zone**, emphasized proving the value of interventions to parents and caregivers: "They don't ask for an RCT before putting their kids in after-school programs," he says. Tools such as participant councils and town halls shift power to participants by asking them to interrogate data and prioritize near-term action.
- 2. Who determines what impact matters most?** Bias is inherent in all measurement and data interpretation, but participatory methods bias in favor of the people who must live with the results. Ben Goodwin, executive director of **Our House**, a multiservice shelter-to-independence provider in Arkansas, began surveying clients in 2015 with a survey tool called **Listen4Good**, developed by Shared Insight. Goodwin now gives data from ongoing biannual surveys to a client council to

interpret and distill into an action agenda: “We’re trying to take our own unconscious bias [as managers] out of the picture,” he says. Subarna Mathes, senior strategy and evaluation officer at Ford, cautions: “In every issue, in every field ... we can point to evaluations where claims have been made about impact that did not bear out over time.... [The right] questions weren’t asked because of the biases within the evaluators themselves.”

“In every issue, in every field ... we can point to evaluations where claims have been made about impact that did not bear out over time.... [The right] questions weren’t asked because of the biases within the evaluators themselves.”

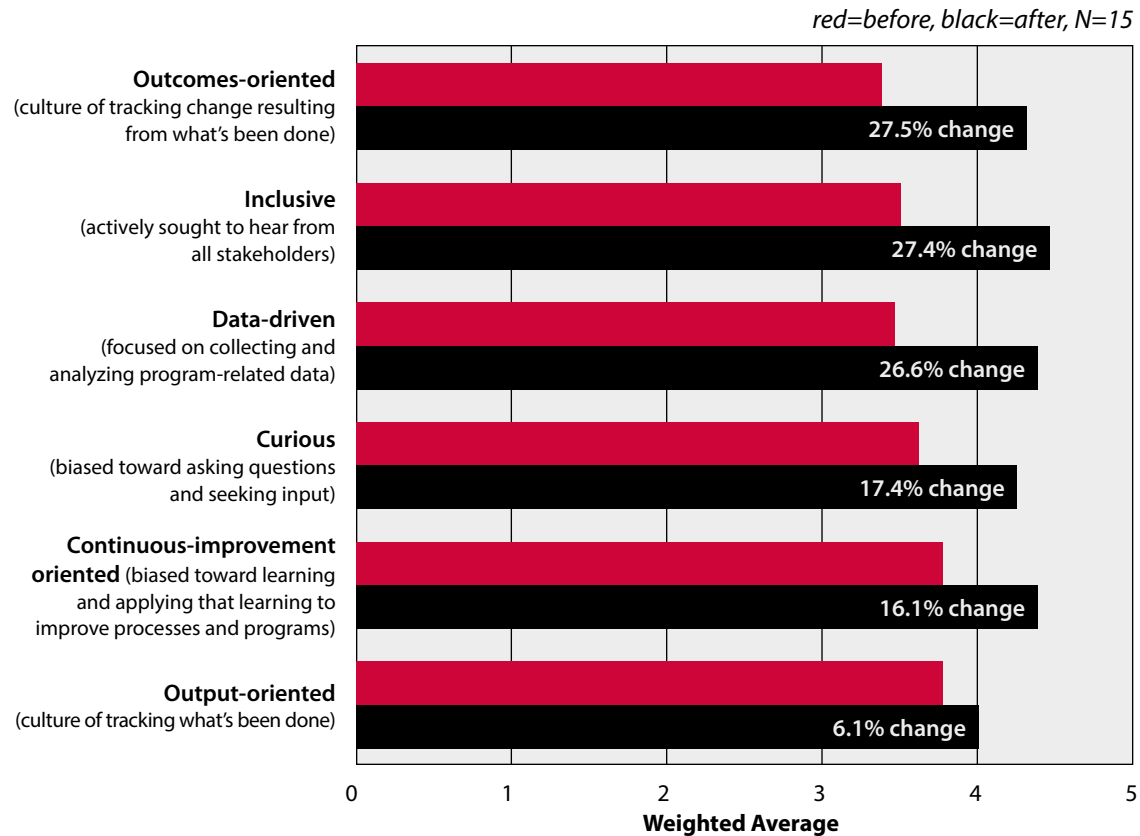
Subarna Mathes, Ford Foundation

3. How do you ask the right questions?

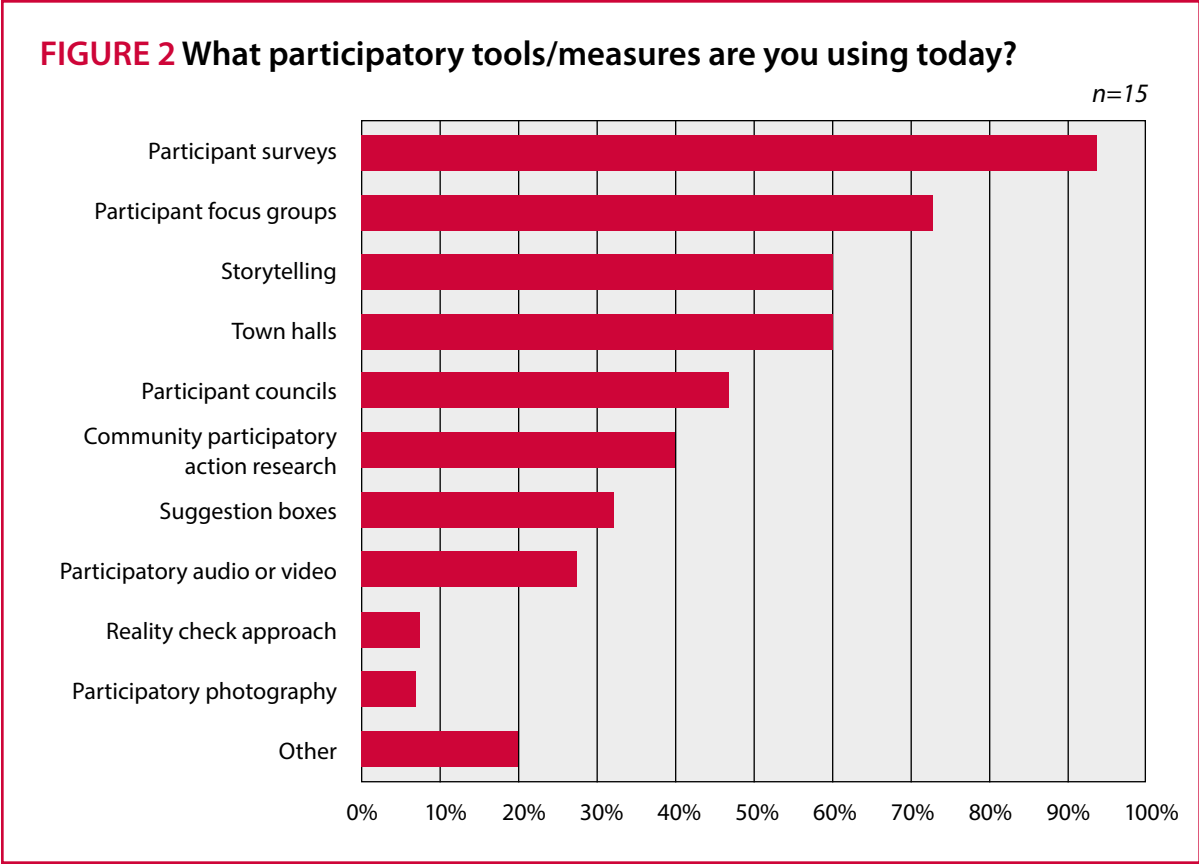
Nonprofits that focus on participant-centered measures tap community wisdom to get those questions right. Outside the U.S., in particular, resource scarcity is fostering innovation. **SOLE Colombia**, based in Bogotá, hosts learning spaces where participants self-organize to address community challenges, like education or safety, and tests questions by sending them out for community feedback. SOLE then listens, learns, and adapts the questions until participants agree the items are ready for a broader dialogue that they feel ready to facilitate. Similarly, **Gift of the Givers**, a disaster relief agency based in South Africa, gathers community members, funders, technical experts, and local authorities for face-to-face, real-time discussions of what’s working and what needs to change. Says Executive Director Muhammad Sooliman: “Everyone has an equal chance to argue for what the community needs, what budgets are required, what local traditions will allow, and, ultimately, what we should do. And then the donors understand why.”

- ### 4. What are the links to program outcomes?
- Several organizations in our focus groups were conducting multiyear research to explore how participant feedback connects to outcomes. Preliminary findings suggested causal links. For example, the **Center for Employment Opportunities**, which helps formerly incarcerated people get jobs and rekindle their lives, indicates that participant feedback revealed complementary components of a program that helped them succeed. “We learned it doesn’t help to have a mentor without a core [job placement] program, but we wouldn’t have learned that through purely quantitative outcome evaluation,” says Ahmed Whitt, director of learning and evaluation. Pace also uncovered links to outcomes by conducting statistical analyses (correlations, regressions, ANOVAs, *t*-tests, and structural equation models) of girls’ perceptual responses in its Listen4Good surveys and empirical data, such as school attendance, grades, and interactions with juvenile justice. To date, Pace has found the strongest link between teacher retention and girls’ feeling more respected and staying longer in the program, with tenure in the program statistically proven to positively influence girls’ results. Nurse-Family Partnership (NFP), a national maternal-child health program known for rigorous measurement of clinical outcomes in healthy births and babies, used feedback to identify program elements that enhanced retention of young mothers in the program.

FIGURE 1 On a scale of 1 (low) to 5 (high), please rate how embedded the following characteristics in your organizational culture were before and after implementing participatory approaches



5. **What are the wins for society?** Perhaps the most forward-facing results of participatory research are advocacy wins that change systems and society. **Think of Us**, a service-delivery and advocacy group for foster youth, hires young people who have experienced foster care to be members of its research teams. When Think of Us recently reported out testimonials from foster youth about their experiences in group homes, the insights prompted several states to voice support for policy changes within hours. Pace, too, uses direct input from its girls to identify local, state, and federal policies that need reform—and the community members who must be involved. One of Pace’s successful community campaigns lobbied for misdemeanor and civil-citation legislation so that law enforcement could censure girls for petty crimes without arresting them. Florida has also funded other prevention measures to keep girls out of the juvenile justice system. As a result, over the past decade, the number of girls arrested annually in Florida has dropped by about 65%.



6. **What’s the impact of the method itself?** Meanwhile, nonprofits like Think of Us are shifting power and helping participants economically by hiring and paying them to conduct research. **Lake County Build a Generation** in Leadville, Colorado, hires constituents to interview peers on local issues ranging from boosting kids’ interest in outdoor activities to expanding affordable housing. Executive Director Katie Baldassar says that Build a Generation links research findings with outcomes by correlating survey results with evidence-based practices: “We can show that the average amount of time Lake County youth spend in nature increased by an hour a week [with a redesign of school and out-of-school time practices], and there is evidence that being in nature improves youth health.”

When our focus-group participants ranked a series of characteristics of their nonprofits before and after the organizations implemented participatory methods, they noted their work became more outcomes-focused (27.5% higher on average), more inclusive (27.4% higher), and more data-driven (26.5% higher; see **Figure 1**). More than 70% of the organizations we studied reported that participatory methods (most often surveys, focus groups, storytelling, and town halls; see **Figure 2**) helped them *define* relevant outcomes, whereas experimental methods often framed an outcome so narrowly that the findings lost relevance. Says Culliton, of Boston Uncornered: “One of the reasons we didn’t do an RCT is because [the research organization] wanted to test ‘Does paying people work?’ Clearly, we’re all being paid to do our jobs. So, yes, it works!”

“The most forward-facing results of participatory research are advocacy wins for system changes and society.”

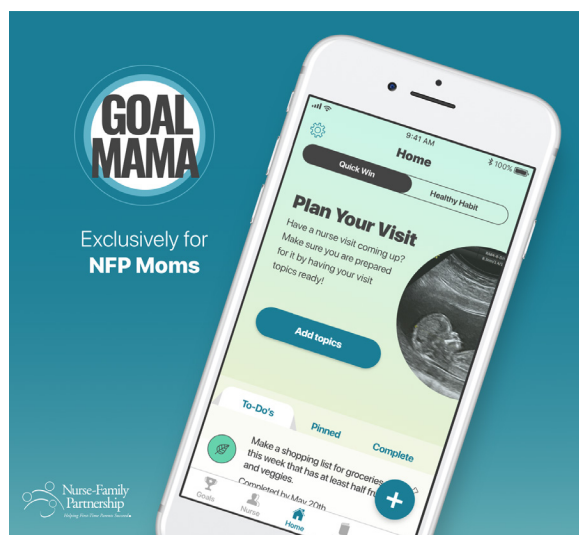
Embedding Equity in Organizations

Nonprofits that focus on participatory evaluation have noted positive knock-on effects in other areas of their work: they are generating more equitable impact inside their organizations, spurring more-inclusive practices in hiring, service delivery, talent development, technology, and communications. Indeed, their shift in values—regarding who is perceived as an expert—permeates myriad organization-building decisions.

Hiring. Sixty percent of our study respondents said the use of participatory approaches in evaluation led them to hire differently—specifically, to prioritize candidates who experienced the issues their program aimed to solve, were good listeners, respected the community, and were participatory-tool savvy. And *all* respondents reported increasing communications with program participants, often by SMS or text. These shifts in capacity bolstered organizational adaptability and development of workplace tools that facilitate taking stock—and implementing change—quickly.

As part of its measurement pivot, Pace, supported by an investment from the [Edna McConnell Clark Foundation](#), or EMCF (which also funded Pace’s original RCT), built an internal measurement team with a Salesforce.org tracking system to correlate participant feedback with empirical measures such as school attendance and juvenile justice involvement. Pace also reinforced functions that fostered participation, greatly expanded its IT team, and invested in technology that improved connectivity among sites, participants, and staff members’ homes. Ultimately, by blending participatory and empirical research, Pace shifted from a compliance-driven partner of the juvenile justice system to a future-focused agent of change with a practical goal—of developing socially, emotionally, and physically healthy, educated, and successful girls—that permeated all of its departments.

Pace also developed internal processes that embraced the same feedback culture it had elicited in its programming. It created girls’ leadership councils at every site and a statewide council representing each site. The latter help design, execute, and interpret program research and evaluation; conduct focus groups with peers; aid in interviewing new hires; and contribute to program decisions. Their



Nurse-Family Partnership learned it could boost participants’ satisfaction by making materials accessible via mobile phones and created an app called Goal Mama. (Photo credit: Nurse-Family Partnership)

deep engagement motivates staff: team-member turnover has declined by nearly two-thirds within five years, and productivity and engagement have increased by more than a quarter.

Service delivery. After starting to gather participant feedback five years ago, **Nurse-Family Partnership** learned it could boost moms' satisfaction with the program by making materials and resources accessible via mobile devices. NFP service delivery underwent a **digital transformation**—from in-person meetings between moms and nurses with binder-based records to an exclusive, goal-setting platform called **Goal Mama**, just for NFP moms, which lets the moms communicate with their nurses between visits and save tips and notes on the digital app. Quality coordinator Alexa Cares became NFP's first feedback manager and started getting direct input from moms. In tandem, the nonprofit amended its **values** to include "actively listen to ... our families." As the feedback function is expanding to 90 sites, NFP revised its survey and data-capture software, shifting from generating reports manually to using Qualtrics, which automatically aggregates survey responses so that sites can use the data to make timely decisions.

More than 60% of our survey respondents said their tech know-how and cross-sector relationships, grown through participatory practices, helped their response during the COVID-19 pandemic. For example, with social distancing restrictions, NFP nurses rapidly shifted to communicating with their clients by phone or videoconference. Meanwhile, the organization partnered with Verizon to provide equipment to first-time moms who lacked phones, and logistics company Action Technologies Group delivered them pro bono to nurses to distribute. Not all went smoothly. "The pandemic presented significant distribution and logistical challenges in getting phones into our clients' hands," says Ashley Stoneburner, NFP senior research analyst. Despite obstacles, the cross-sector collaboration allowed NFP to reach 3,800 clients, enabling ongoing connection with their nurses throughout the pandemic.

“Sixty percent of our study respondents said the use of participatory approaches in evaluation had led them to hire differently—specifically, to prioritize candidates who had experienced the issues their program aimed to solve ...”

Talent development. Our focus-group members also created roles for participants on their teams so that they could build skills and become eligible for permanent positions. Think of Us creates positions for foster youth to be junior user-experience researchers. Says Senior Director Sarah Sullivan, "They get to learn from us how to be a UX researcher, and we get to learn from them." NFP has staffed its **outreach team** with alumni who share their own experiences to motivate new participants to join. At Boston Uncornered, participants receive \$25,000 annual stipends during their training, and 50% of staff are former participants. **My Life My Choice**, a Boston-based mentoring and prevention nonprofit for sex-trafficked youth, is *led* by survivors of this experience.

Nonprofits need to rethink recruiting to persuade their alumni to apply for such research roles. At Build a Generation, 25% of community connectors are people who have direct experience with challenges such as access to food or affordable housing, and the budget line for them keeps growing. "These are not just extra contractors; this is core to what we do," says Baldassar. A lot of times, however, she hears

that potential applicants don't see themselves as eligible. "What we're really trying to think about is how to phrase job descriptions so that the people we serve see themselves in the job."

Technology and communications. For our international focus-group participants, technology and tech-savvy talent have long been capacity builders and cost savers, using tools like WhatsApp and Facebook. Doussouba Konaté, global program manager of nonprofit **Accountability Lab**, in Mali, works to ensure that community members, particularly in remote areas, are consulted by those in power about the decisions that affect their lives, building accountability among citizens, governments, the media, and the private sector. "We work with volunteers called Community Frontline Associates (CFAs) to gather information," says Konaté, "with one CFA capturing the opinions of 50 to 100 people per questionnaire." For utility and cost-effectiveness, CFAs use KoBoToolbox, an open-source app that aggregates data from interviews and has a voice-entry option. In one study of people with disabilities (many with low or no literacy), the CFA gave participants phones with KoBo and asked them to record diary entries of their experiences over a few months, tracking when, where, and how they faced discrimination. Accountability Lab shares such testimonies through local radio, music, and films; local language bulletins; and gatherings with local leaders to clarify concerns, forge solutions, and make public plans and timelines for implementing them.

Sooliman says Gift of the Givers embeds journalists in its relief teams to write eyewitness stories of interventions and their results, and equips staff with camera phones and recording devices for field research. He also has built counseling and training capacity to help staff sensitively communicate with disaster victims as they gather feedback. In addition, Gift of the Givers trains staff in conflict resolution in disaster-struck communities. Using feedback from the field, the organization has refocused volunteer recruitment from seeking technical experts to identifying collaborators and communicators.



Accountability Lab community frontline associate in Mali's neighboring country Burkina Faso captures survey responses from a community member. (Photo credit: Accountability Lab)

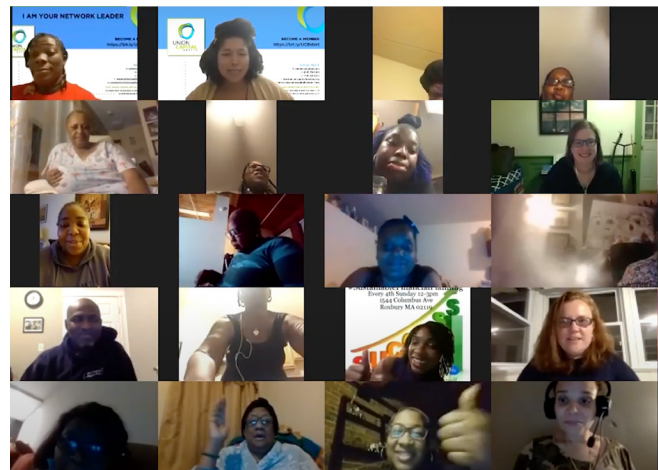
Growing Resilient Cultures

Nonprofits' adaptability was tested mightily by the COVID-19 pandemic and racial reckoning following George Floyd's murder, with Charities Aid Foundation America finding that 71% had to adapt programs. Those in our focus groups that had built capacity for participatory evaluation found that their cultures supported quick, smart shifts to new and virtual ways of listening to, and working with, participants to respond to trauma and injustice.

Connecting. "When the pandemic hit, so much was changing and unknown," says Goodwin, of Our House. "Our staff just naturally went to: 'Hey, let's send out a survey. Let's see how many participants are thinking they might put their kids in in-person school versus virtual.'" Thirteen of the 15 nonprofits we studied said their focus on participatory methods made them more resilient in navigating the pandemic, and all 15 said they benefited from having developed these virtual, rapid ways to learn participants' thoughts and needs.

“[Nonprofits] that had built capacity for participatory evaluation found that their cultures supported quick, smart shifts to new and virtual ways of listening to, and working with, participants to respond to trauma and injustice.”

Union Capital Boston (UCB), for example, which helps build social capital among low-income residents by rewarding community involvement with Visa gift cards, learned several years ago through biannual surveys that members wanted to meet one another. So, UCB started hosting weekly Network Nights, where members gathered in a physical location to check in, lead discussions on topics that they nominate, and exchange favors and goods. These mechanisms were critical in 2020. "At the very start of the pandemic, we were able to raise \$400,000 for COVID relief and distribute it quickly via \$150 gift cards, because we had the rewards model," says Jalina Suggs, director of networks. "And we converted our Network Nights to virtual forums."



With the outbreak of COVID-19, Union Capital Boston quickly shifted its weekly, in-person Network Nights (left), where community members proposed and discussed issues and exchanged favors, to virtual gatherings (right).

(Photo credit: Union Capital Boston)

After George Floyd’s murder in Minneapolis, Network Nights became even more crucial. “Our next Network Night was an opportunity for members to share how they were experiencing it,” says Suggs, who invited organizations working on restorative justice and defunding the police to attend. “Processing racism has been ongoing since, being explicit that UCB is an anti-racist organization led and founded by a white man,” says Suggs.

Conversing. Building a culture of feedback also led to important staff conversations during the crises of 2020. After Floyd’s murder, staff at My Life My Choice worked with leadership to review policies and practices, and identify and address any that could, unintentionally, fail to empower its diverse and survivor-led staff. “We had important and very hard conversations in that period to build greater racial equity into our culture,” says AJ Espensen, the nonprofit’s manager of outcomes and evaluation. During that same period, Boston Uncornered had staff layoffs, prompting an unsuccessful bid to unionize and complaints about CEO Culliton, who is white, making decisions. Culliton says, “The choices can seem negative at first, but really it’s the process of taking ownership ... to create change we hope to see.” Cofounder Michelle Caldeira adds: “It was simpler, too, to put that decision on Mark and ignore the fact that there was also a Black woman leading the organization! There has been a ton of continuing to build trust and allowing all of us to exhibit our co-ownership in the work.” (Espensen’s, Culliton’s, and Caldeira’s observations about a push for participant feedback spurring tough-but-important staff conversations and ownership reflect our broader focus-group findings.)

Deploying resources. Investments in resources, like technology, to implement participatory measures allowed nonprofits to continue to reach participants during the crises. Pace, for example, extended its technology during the pandemic from supporting 527 staff to an additional 2,000 girls—and implemented



Pace extended its technology during the pandemic from supporting 527 staff to an additional 2,000 girls—including providing internet-enabled tablets, laptops, and Microsoft teams accounts—and implemented remote services with feedback channels. (Photo credit: Pace Center for Girls)

remote services with feedback channels for the girls. As a result, Pace engaged with over 90% of its girls, all of whom qualify for free or reduced lunch in school, at least once per week during school closures, and with 75% seven or more times per week—delivering food, computers, and telecounseling. Ultimately, 91% of girls completed the program in 2020 (vs. 81% in 2019), and 88% improved academically in the fourth quarter (vs. 70% the prior year). Meanwhile, [EdWeek](#) found that in one out of three high-poverty communities across the U.S. (those with 75% or more students on free or reduced lunch), students had no engagement with schools at the outset of the pandemic.⁵

Deepening virtual engagement during the COVID crisis made it clear that budget lines for extending technology to participants should be baked into post-pandemic budgets. Suggs, of UCB, for example, which has been paying for a high-functionality Zoom account its members can use to host their own events, doesn't anticipate eradicating virtual Network Nights, as they have helped many who have chronic transportation challenges to become involved, and at a lower cost than in-person meetings. Indeed, almost all the nonprofits we studied found that their investments in people and practices to implement more-equitable evaluations had strengthened their reflexes to respond to crises.

A Holistic Approach to Build Equitable Evidence

The nonprofit push to build evidence, capacity, and culture more equitably is part of a nascent ecosystem movement. In the past six years, funders like the [10 foundations](#) that collaborate as core investors in Shared Insight have given grants to help nonprofits develop tools for listening to their participants. Many of these foundations also participate with Ford in a working group to shift power from largely white evaluation staff to the participants themselves—and to diversify talent pipelines for evaluators. In the past few years, equity-focused evaluation intermediaries have sprung up; these include [We All Count](#), committed to [bringing equity to data science](#); the [Equitable Evaluation Initiative](#), offering an [Equitable Evidence Framework](#) for shifting practice; and Project Evident, which is helping nonprofits derive learning agendas from the questions of their participants and social sector leaders. Even big evaluation shops that have reputations for large-sample, experimental, or quasi-experimental methods are shifting to holistic approaches.

Implementation. Laird, of RTI, describes a blend she implemented for an impact assessment of a California social enterprise that had hired returnees from incarceration to train and work in its shops before entering the competitive job market. “The study called for using an RCT and a quasi-experimental design to test the impact of working at a social enterprise on the self-sufficiency of participants

18 months later, including being employed,” says Laird. “Then Shared Insight approached us and asked, ‘What would you think about embedding a perceptual feedback study in the analysis?’” The idea was to see whether feedback data could be an early proxy for subsequent outcomes, a less costly option for organizations that couldn't afford an RCT. Laird added in the feedback component as a simple survey and then tested correlations to see if the data on participants' perceptions a third of the way through the program could predict whether, 18 months later, they had found a livelihood. “After doing

“The nonprofit push to build evidence, capacity, and culture more equitably is part of a nascent ecosystem movement.”

factor analysis, we found associations,” says Laird, “with a proximal outcome of exiting the program positively—to either a competitive job or school.” When she modeled them, two associations rose to the top. One was a confidence measure: “If participants were fearful about leaving their supportive work environment, they didn’t fare as well,” according to Laird. The second, similar to causal links identified at Pace, showed that having a strong connection to a staff member boosted successful job placement. RTI is now completing the analyses to examine whether those same factors predict having a job 18 months after entering the program.

With these insights, the social enterprise could take action to improve its outcomes. To address the confidence gap, the organization brought back alumni to speak to participants’ fears, share their transition experiences, and explain how the training had prepared them. For strengthening connections to staff, it looked to build staff retention and strong interactions with job coaches. “There was total value in both [the RCT and the perceptual feedback],” says Laird.

Laird observes that funders are increasingly embracing holistic impact assessment. The Edna McConnell Clark Foundation, which a decade ago pioneered aggregating funds to make long-term, multimillion-dollar investments in nonprofits that were ready to grow their impact, prioritized the results of RCT studies as an important indicator of such readiness. EMCF went on to become a core funder of Shared Insight. Meanwhile, its former director of evaluation and learning, Gabriel Rhoads, became a founding team member of Project Evident, where he is now managing director of evaluation services. “Outcomes still matter,” says Rhoads, “but there has been an equity shift to rebalance power. We now ask, ‘What do social sector leaders want to know? What do participants want to know? Who is asking? Who gets to look at the data? Who gets to decide?’” Mathes, at Ford, adds: “I have not engaged with the RCT community because, for most of Ford’s work, it won’t get us the insights we need. You can’t ignore the ethical piece: denying service. And we can’t explain away the need for some degree of participant participation.”

Pace’s own blend of tools includes its Girls Coordinating Councils (GCCs), where community stakeholders, including girls themselves, are given the space to influence favorable conditions for girls’ and young

You’ve been arrested...	You’ve been arrested, now what?!	Important Reminders	If your child has been arrested, here is what you need to know...
<p style="text-align: center;">You are arrested, now what?</p> <p style="text-align: center;">If you are arrested or received a Notice to appear in court, here is what you need to know...</p>	<p style="text-align: center;">If arrested, am I guilty? No, being arrested does not mean you are guilty, but missing court will result in an arrest warrant.</p> <p style="text-align: center;">Why should I go to court? Attending your court date allows you to tell the judge your side of the story and ask questions.</p> <p style="text-align: center;">How should I prepare for court? Check your court date and time, arrange reliable transportation, look presentable, arrive early, use an inside voice, stay calm, listen to the judge, and ask questions if you don’t understand something.</p> <p style="text-align: center;"><i>Legal Aid Service of Broward County LAW-line is available for youth and families who have questions or concerns about youth’s involvement with the delinquency system.</i></p> <p style="text-align: center;">954-318-4531</p>	<p style="text-align: center;">Court Date:</p> <hr/> <p style="text-align: center;">Time:</p> <hr/> <p style="text-align: center;">Judge’s Name:</p> <hr/> <p style="text-align: center;">Courthouse:</p> <hr/> <p style="text-align: center;">Courtroom Number:</p>	<p style="text-align: center;">If my child is arrested, are they guilty? No, being arrested does not mean your child is guilty, but missing court will result in an arrest warrant for your child that has added financial cost to you as a parent/guardian.</p> <p style="text-align: center;">Do I need to go to court with my child? It would be best if you were at court with your child, but your child must attend.</p> <p style="text-align: center;">How should I prepare for court? Check your child’s court date and time, call in your child’s absence from school, arrange reliable transportation, be prepared for parking costs, and arrive at least 30 minutes early.</p> <p style="text-align: center;"><i>Legal Aid Service of Broward County LAW-line available for youth and families who have questions or concerns about youth’s involvement with the delinquency system.</i></p> <p style="text-align: center;">954-318-4531</p>

Pace’s Girls Coordinating Councils created resource cards for girls to carry in event of arrest. The next year arrests declined 16% and failures to appear in court 27%.

women's healthy development in their communities. In 2018, a GCC in Broward County, Florida, tackled the county's rate of detaining girls for failing to appear at their court hearings. When the girls interviewed judges, probation officers, and youth, they found that often girls did not understand the consequences of an arrest, and those who had been arrested forgot their hearing dates or struggled to find transportation to the court. The girls' research team created a video with avatars portraying what arrested girls could expect upon entering the juvenile justice system. They also created cards in the languages spoken in their communities, which teenagers could carry with their cellphones in the event of arrest, with hotline numbers for case managers, transit information, and ways to contact the court in the event of a delay. The following year, arrests in Broward declined 16%, and instances of failing to appear after arrest dropped 27%.

Expansion. At the same time, the design of RCTs themselves has evolved to embrace broader objectives and more-nimble execution—and to seek out experimental conditions that make the most of naturally occurring control groups. As the Biden administration seeks to introduce universal preschool, one such RCT, which tested life outcomes for Boston children who attended preschool, has become a case in point.

Study findings, released in May 2021 by the School Effectiveness and Inequality Initiative (SEII), used pre-K admissions lotteries to estimate the effects of large-scale public preschool in Boston on college attendance and preparation, standardized test scores, and behavioral outcomes. The lotteries, instituted because Boston didn't have a budget to offer preschool to all, meant that thousands of otherwise similar children would have different life experiences based on random chance, and SEII economists from University of Chicago, MIT, and UC Berkeley were able to use this naturally occurring control group to design an RCT. They also designed the study to test beyond standard measures like state achievement tests.

Ultimately, the RCT found that preschool enrollment boosted rates of high school graduation, college attendance, and SAT test-taking—in the latter two cases, more for boys than girls. And the study found that preschool reduced disciplinary measures such as suspensions and juvenile incarceration, again with greater impact on boys than girls. However, the study uncovered no detectable effect on the most standard measure tested—state achievement test scores—nor differences by income or race.⁶ These findings illustrate not only the potential for preschool to mitigate inequality but also the importance of measuring long-term and non-test-score outcomes, and of segmenting results. And they raise the question: What would SEII have learned about other enabling conditions if economists had included participant feedback in the study?

My Life My Choice (MLMC) applied mixed measures when, in 2015, it engaged university researchers to evaluate how well its 10-session Exploitation Prevention Curriculum helped prevent the commercial sexual exploitation (CSE) of at-risk girls. The ensuing longitudinal, multisite, three-year evaluation combined biannual surveys of girls taking the

“Beyond making progress in blending in participatory measures, randomized evaluation tools have become more agile and less costly ... Today, participants can opt in to online studies rather than being denied service, all of which lowers barriers to combining evaluation methods holistically.”

curriculum (to understand their knowledge of, attitudes toward, and access to trusted help and information about CSE), with empirical measures of changes in CSE (quantifying sexually explicit behavior, dating abuse, and alcohol and marijuana use). A multivariable-adjusted model showed that 100% of girls in the program increased their knowledge of CSE and its risks, halved their experience of CSE, and were twice as likely to report dating abuse versus at-risk girls not in the program. Moreover, youth perceived their facilitators as honest, warm, and trustworthy, validating not only the curriculum but also its staffing, something empirical research alone could not have ascertained. The study also applied mixed measures to test the effectiveness of MLMC's Survivor Mentor Program, finding mentored youth were three times less likely to report having been commercially sexually exploited, 54% less likely to be arrested or detained by the police, and 1.4 times more likely to have adequate social support, the latter discovered thanks to participatory methods.

Beyond making progress in blending in participatory measures, randomized evaluation tools have become more agile and less costly, with growth in social media and machine learning making big data more accessible. Today, participants can opt in to online studies rather than being denied service, all of which lowers barriers to combining evaluation methods holistically. Susan Athey, the Economics of Technology Professor at Stanford University, worked on a study on stemming the spread of COVID misinformation on Facebook in Nigeria and Kenya, which involved recruiting participants through Facebook ads and using chatbots on Facebook Messenger to survey and expose participants to treatments targeted at combating the misinformation. The [study](#) involved approximately 10,000 respondents at a total cost of about \$6,000.⁷

Next Steps for Funders, Evaluators, and Nonprofits

With the expansion of experimental and participatory approaches comes greater opportunity to mix methods. And the opportunity comes at an inflection point in building evidence of social impact, one that Mathes at Ford describes as a "transitional moment in the field of evaluation where what actually counts as truth, what counts as valid, what counts as data is shifting and needs to shift." We can't generalize outcomes from a single longitudinal study, but we must learn under what conditions an intervention can work by seeking participants' insights at every step of evaluation, design, and delivery, always remaining aware of power and context. It remains unclear whether and when this message will get through to public funders, who look to RCT proof points for decisions to invest tax dollars in expanding social programs.

For now, as funders (public and private), evaluators, and nonprofits proceed on this journey, they all should ask to what extent they can affirm the following:

1. Do we have a process for connecting with and bringing the voices of those most affected by our interventions to influence our approach to evaluating impact?
2. Are we disaggregating the data we gather by race, age, and any other relevant criteria to understand the disparate experiences of groups of participants within the overall results?
3. Are we approaching our evidence building through a racial equity and inclusion lens by identifying and testing questions through appropriate focus groups or panels of participants?

4. Are we calibrating outcomes to ensure they are equitable and not determined or predictable by other innate factors (e.g., gender, disability, or race)?
5. Are we using the data to inform programs and strategies that are themselves in the service of equity?

Entities making intentional and ongoing progress toward affirming all five questions are well on their way to building more-equitable evidence.

Lymari Benitez is senior director of program information and impact at Pace Center for Girls. Yessica Cancel is Pace's chief operating officer, and Mary Marx its chief executive officer. Katie Smith Milway is principal of MilwayPLUS and a senior advisor at The Bridgespan Group. The authors thank reviewers Mariah Collins, Lindsay Louie, Subarna Mathes, Rick Moyers, Gabriel Rhoads, Kyle Rinne-Meyers, Lissette Rodriguez, Dustin Sposato, Teddy Thompson, Melinda Tuan, and Fay Twersky for their contributions to this article.

Endnotes

- 1 Fabienne Doucet, “Identifying and Testing Strategies to Improve the Use of Antiracist Research Evidence through Critical Race Lenses,” The William T. Grant Foundation, The Digest, Issue 6, January 21, 2021.
- 2 Sociologist W. E. B. Du Bois in 1898 used **community surveys** to understand structural racism. Sociologists Kurt Lewin, Margot Haas Wormser, and Claire Selltitz in the 1940s and ‘50s used **participatory community self-surveys** to understand individual lived experiences.
- 3 Institute of Development Studies, Key Terms, Participatory Methods, <https://www.participatorymethods.org/glossary/participatory-action-research>.
- 4 Institute of Development Studies, “Participatory Monitoring and Evaluation: Learning from Change,” IDS Policy Briefing, Issue 12, November 1998.
- 5 Stephen Sawchuck and Christina A. Samuels, “Where Are They? Students Go Missing in Shift to Remote Classes,” Education Week, April 10, 2020.
- 6 Guthrie Gray-Lobe, Parag Pathak, and Christopher Walters, “The Long-Term Effects of Universal Preschool in Boston,” School Effectiveness and Inequality Initiative, discussion paper, May 2021.
- 7 Molly Offer-Westort, Leah Rosenzweig, and Susan Athey, “Optimal Policies to Battle the Coronavirus ‘Infodemic’ among Social Media Users in Sub-Saharan Africa,” **revised preanalysis plan**, February 25, 2021. See also: Susan Athey, “How Nonprofits Can Leverage Randomized Evaluation Tools and Measure Social Impact Effectively,” SSIR.org webinar, February 24, 2021.



6745 PHILIPS INDUSTRIAL BLVD.
JACKSONVILLE, FL 32256
PACECENTER.ORG



*This work is licensed under the Creative Commons
BY-NC-SA 4.0 license. To view its terms visit
<https://creativecommons.org/licenses/by-nc-sa/4.0/>*

katiesmith.milway@gmail.com
www.milwayplus.com
781-244-5005