Philanthropy, Outcomes, & Impact

Mark Elliott
Economic Mobility Corporation

November 2019

MOBILITY

The Economic Mobility Corporation (Mobility) identifies, develops, and evaluates programs and policies that enable disadvantaged individuals to acquire the education, skills, and networks needed to succeed in the labor market so that they can support themselves and their families.

Mail:
Economic Mobility Corporation, Inc.
50 Broadway, Suite 1604
New York, NY 10004
Telephone:
212.280.6975
Email:
info@economicmobilitycorp.org

Acknowledgements

Much appreciation to Caitlyn Brazill, Plinio Ayala, Jerry Rubin and Sister Pearl Ceasar for sharing their thoughts on this subject and for making many helpful suggestions on the piece. Harry Holzer, Jon Baron, Russell Pomeranz and Brad Dudding read the essay and offered their comments and encouragement as well. And my colleague Anne Roder oversaw the analysis and did her best to ensure I didn't push the argument too far.

The figures used in this brief are based on our report, *Nine Year Gains*, which was supported by Arnold Ventures. The analysis for that report was conducted for Mobility by the Ray Marshall Center at the University of Texas.

Caitlin van Duesen copyedited the brief and Penelope Malish developed the design.

In 2012, the Bridgespan Group, a nonprofit organization based in Boston, published a report titled “Building Capacity to Measure and Manage Performance.” The report chronicled the growing movement in the nonprofit sector to incorporate an outcomes-oriented culture related to each organization’s mission. The report begins: “Over the last decade, the social sector has witnessed a dramatic shift towards a results focus, with nonprofits and other mission-driven organizations feeling immense pressure from government and philanthropy to demonstrate their impact.…Nonprofits that want to have more impact in the world also need the capacity to use measurement to improve their performance.”[1]

In our experience, this pressure has only increased since 2012. In recent years, we’ve seen the rise of “impact investing,” new public sector funds focused explicitly on outcomes, such as the Social Innovation Fund, and the creation of financing tools such as social impact bonds and Pay for Success projects.

Foundations have become explicit about their reliance on evidence in grant-making:

- “We use rigorous metrics and data to take bold action and create impactful solutions, we support the work of the very best nonprofits providing direct-service work in the poverty fight, and we follow the data wherever it leads us.”[2]
- “We advance research and solutions to overcome the barriers to success, help communities demonstrate what works and influence decision makers to invest in strategies based on solid evidence.”
- “We focus on community-led initiatives, partnerships with all levels of government, stronger nonprofits, and the use of data to track progress and measure improvement.”

This focus on evidence has become pronounced in the workforce development field, where rigorous evaluations of programs have generally found little or, more often, no impact.[3] Private foundations have fueled the push for evidence of effectiveness by requiring detailed outcomes-oriented reporting from grantees. Not surprisingly, a bevy of consultants as well as for-profit and nonprofit organizations have sprung up, touting their ability to work with organizations to analyze and improve their performance. Indeed, Mobility has worked with several nonprofit workforce development organizations to analyze their data and provide recommendations about ways to improve outcomes.

Focusing on participant outcomes to improve performance is certainly better than not focusing on outcomes at all. But in our work—including conducting randomized controlled trial evaluations of several leading organizations that have shown large, statistically significant earnings impacts—coming to sound conclusions about what constitutes strong performance is not as simple as it might seem.

In this essay, we focus on two concerns about the outcomes movement:

- the enormous burden imposed on organizations to collect and report data to funders, often without compensation
- how pressure to improve performance through a focus on outcomes may actually diminish an organization's true impact on the people it serves

Daunting data collection and reporting requirements

As foundations have become more outcomes-oriented, they are asking their grantees to do more and more data collection and reporting. Each foundation typically has its own reporting requirements, but few recognize that the scale of philanthropy's collective reporting demands presents a difficult challenge for organizations. Staff at two leading workforce development organizations—Per Scholas and JVS Boston—shared their observations about working with foundations on data collection and reporting. Both organizations have demonstrated large earnings impacts for their participants in rigorous, randomized controlled trial evaluations. If foundations were to trust any organization's performance, these two would certainly qualify.

Per Scholas is a nonprofit workforce development organization headquartered in the South Bronx, with programs operating in several cities across the country. Founded in 1995, Per Scholas prepares low-income job seekers for careers in information technology. It is currently supported by more than 150 foundations, to which it submits nearly 200 reports annually. In addition, it manages approximately 30 government contracts. The level of reporting required by these funders varies enormously: some require quarterly updates, while others simply ask for a narrative report as part of a renewal proposal. According to Caitlyn Brazill, Per Scholas's executive vice president, sometimes the foundations that provide the smallest grants are the most demanding: "There is little relationship between the size of the grant and the reporting frequency or level of detail requested."

Many foundations have their own interests for which they require Per Scholas to collect and report data. One is interested in outcomes for veterans; another wants to know the effect on poverty alleviation, which requires staff to ask program applicants details about the income of every member of their household. Per Scholas has invested heavily in developing and maintaining a management information system built in large part to fulfill foundation reporting requirements.

JVS Boston was founded in 1938 to serve Jewish immigrants and refugees seeking to join the American workforce. Annually, it serves over 15,000 people who are seeking to sharpen their job skills, advance their education, and secure employment. It is supported by 33 foundations, 29 government contracts, and 30 direct contracts with employers. Jerry Rubin, president and CEO, says: "We're often asked by funders to provide different kinds of data from different sources. And there's very little coordination among them of what's being asked for." He echoes Brazill's sense that there's little relationship between the size of the grant and the reporting requirements it entails. In fact, Rubin believes "there's often an inverse relationship between grant size and data requirements."

It's not clear to staff at Per Scholas and JVS Boston why foundations are asking for so much data. As Rubin says, "You're asking us to collect all this data. What is it for, and does it really help improve your work or ours? This costs real money—which you may not be providing." Plinio Ayala, Per Scholas's executive director, says much of the required data is not helpful to the organization itself: "I can look at eight indicators and know how we're doing."

Jumping to (incorrect) conclusions

When an organization tries to improve program performance, it often analyzes key outcomes by subgroup to determine which groups the organization serves best and which ones seem to do less well. In workforce development, key outcomes typically include program completion and the employment rates and earnings of graduates. Organizations will often compare participants' earnings at the time they enrolled and after they graduated and attribute any gains entirely to the program.

Recommended strategies for improving performance often include serving more people who are likely to achieve the program's employment and earnings goals and fewer of those whom the organization appears to serve less well. This seems simple enough: after all, resources are scarce and should be devoted to people who are most likely to benefit from the program.

But in our view, organizations should not be aiming to achieve the highest performance but the greatest impact: the difference the organization makes in the lives of the people it serves. The most rigorous method of determining whether a workforce development organization has an impact on its participants is by randomly assigning people to a group that receives the organization's services and to a control group, which does not, and comparing the outcomes of the groups. Any statistically significant difference between the two groups can be attributed to the program. This type of evaluation is called a randomized controlled trial evaluation, commonly referred to as the gold standard in program evaluation.

Findings from a randomized controlled trial evaluation that Mobility conducted of Project QUEST offers insight into the distinction between program outcomes and impact. Project QUEST was founded in 1992 through a grassroots effort led by two community organizing groups, COPS (Communities Organized for Public Service) and Metro Alliance. Prompted by the rapid erosion of low-skill manufacturing jobs in the 1980s, QUEST was created to enable residents of San Antonio, Texas, to access good jobs in growing sectors of the local economy by helping them gain skills to meet employers' needs. In its 27 years of operation, QUEST has served over 7,000 people.

QUEST provides comprehensive support and resources to help individuals complete occupational programs at local community colleges, pass certification exams, and obtain jobs in targeted industries.

In 2005, QUEST agreed to take part in a rigorous study to assess its impacts on participants' earnings. The evaluation focused on individuals pursuing training for health-care occupations, including as licensed vocational nurses; registered nurses; medical records coders; and surgical, sonography, and radiology technicians. After QUEST had completed its assessment and interviews, applicants were randomly assigned to the treatment group (who received QUEST services) or to the control group, who did not (although they could still pursue training on their own at San Antonio community colleges and elsewhere). Every person enrolled in both groups was included in the analysis of outcomes. As a result of this randomized controlled trial design, any statistically significant difference in employment and earnings between the two groups can be attributed to QUEST.

In April 2019, Mobility published "Nine Year Gains,"[4] which documents that QUEST had demonstrated the largest sustained earnings impacts ever found in a rigorous evaluation of a US workforce development program. As illustrated in Figure 1, in the ninth year after enrollment in the study, QUEST participants earned $5,239 more than control group members.

Figure 1 **Project QUEST's Impact on Average Annual Earnings during Each Year after Random Assignment (N=410)**

Note: Earnings impacts are reported in current dollars and represent differences in average annual earnings between treatment group and control group members in each year. Statistical significance levels: *** p<.01; ** p<.05; * p<.10

Figure 2 shows what the treatment and control groups' actual average earnings were each year after enrollment. As one can see, in the early years, QUEST participants actually made less than control group members as they pursued occupational training at one of San Antonio's community colleges. Then, beginning in the third year, QUEST participants began making more—and those gains have continued ever since (with some variation year to year). It's notable that control group members' earnings also increased—just not as rapidly as those of the QUEST group.

Figure 2 also illustrates a serious mistake organizations and funders make when they use outcomes such as pre-program and post-program earnings to measure performance. In the year before enrollment, both groups made an average of $11,722. In the ninth year, the QUEST group made an average of $33,644, nearly three times what they made before enrollment. Most organizations would attribute the entire gain to their program. But by tracking the earnings of the control group, which also increased substantially (to $28,404), we can see that it would be incorrect for QUEST to claim credit for participants' entire earnings gains. QUEST's true impact is *the difference between* the earnings of the two groups.

Figure 2 **Average Annual Earnings during the Nine Years after Random Assignment among All Study Participants**

Note: Earnings are reported in current dollars. Statistical significance levels: *** p<.01; ** p<.05; * p<.10

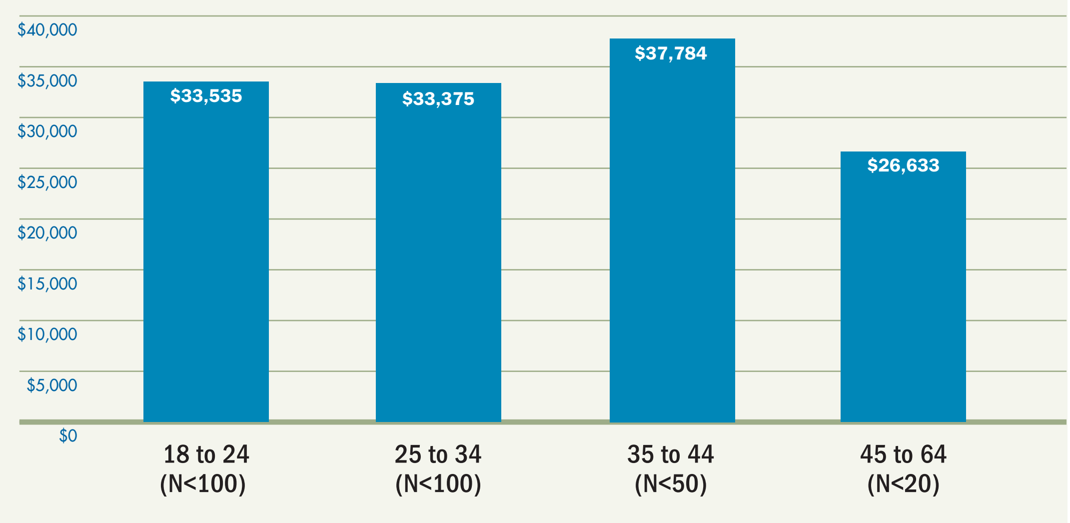
Outcomes versus Impacts as Reflected in Subgroups

By examining the earnings of key subgroups, we can illustrate how easy it is to come to erroneous conclusions about what's driving an organization's performance by focusing solely on outcomes—and, consequently, to pursue strategies that could diminish an organization's impact.[5]

Age

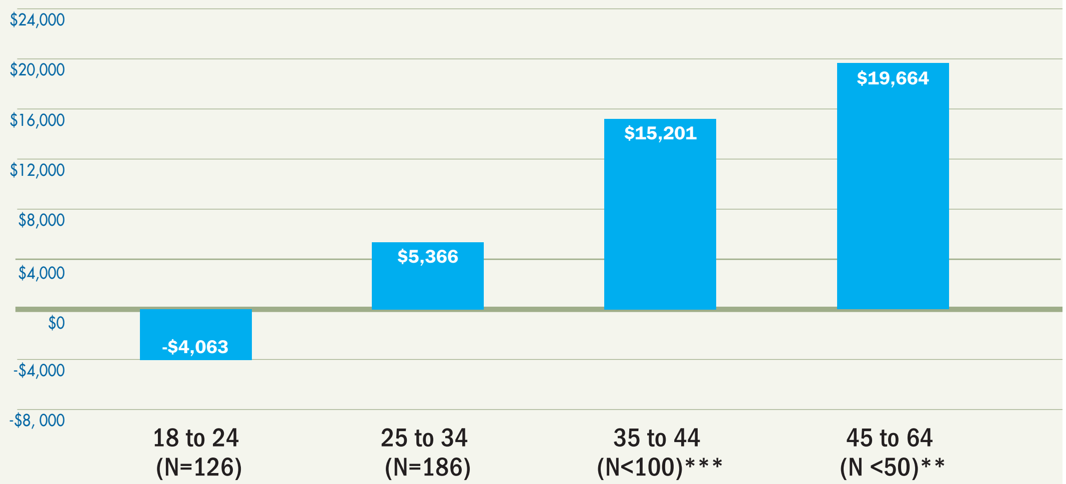
Figure 3 illustrates the average earnings of QUEST participants in the ninth year after random assignment by their age at program enrollment. It's apparent that those who were 45 and older when they enrolled were earning far less in year nine than their younger counterparts. One might conclude that QUEST should stop enrolling older job seekers, or at least enroll fewer of them going forward. After all, in addition to earning less, they have less time remaining to participate in the labor force to reap the rewards of the program's investment in them.

Figure 3 Average Annual Earnings among Project QUEST Participants Nine Years after Random Assignment, by Age at Program Entry



But when we look at the earnings impacts (the difference between the earnings of the QUEST group and the control group) by age, (Figure 4), a very different picture emerges. It becomes clear that QUEST is having the largest impact on the oldest participants: nearly $20,000 in the ninth year after enrollment. Why are the earnings impacts so large? It turns out that many older participants assigned to the control group were, nine years later, no longer working.

Figure 4 **Project QUEST's Impact on Average Annual Earnings in Year Nine, by Age at Program Entry**



Note: Earnings impacts are reported in current dollars. Within group significance levels: ***p<.01, **p<.05
Between group significance levels: ***p<.01 35 to 44 and 45 to 64 vs 18 to 24, **p<.05 25 to 34 vs 18 to 24

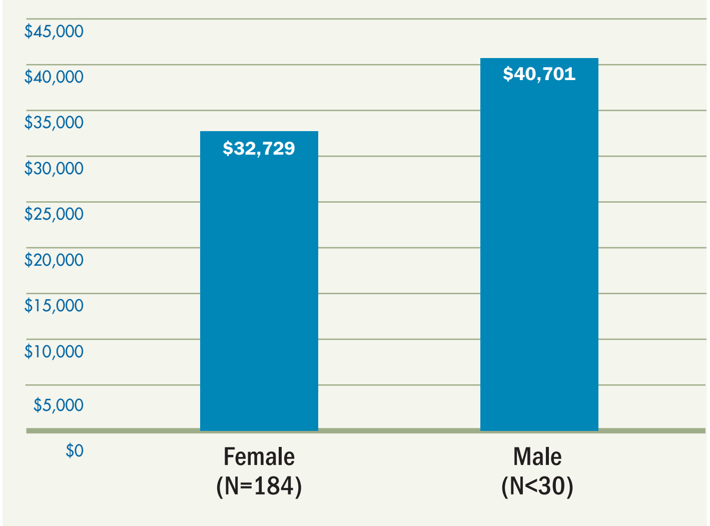
In contrast, in year nine the youngest participants (ages 18 to 24), whom QUEST appeared to be serving reasonably well, based on their incomes, were earning over $4,000 less than control group members.

Sister Pearl Ceasar, who led the community organizing that launched QUEST in the early 1990s and served as its executive director from 2011 to 2017, reflects on these findings: "I would still target recruitment on people in their late twenties and early thirties, but this shows that when you give older workers hope, they are going to be tenacious and become reliable workers. Without that hope, many older workers simply quit."

Gender

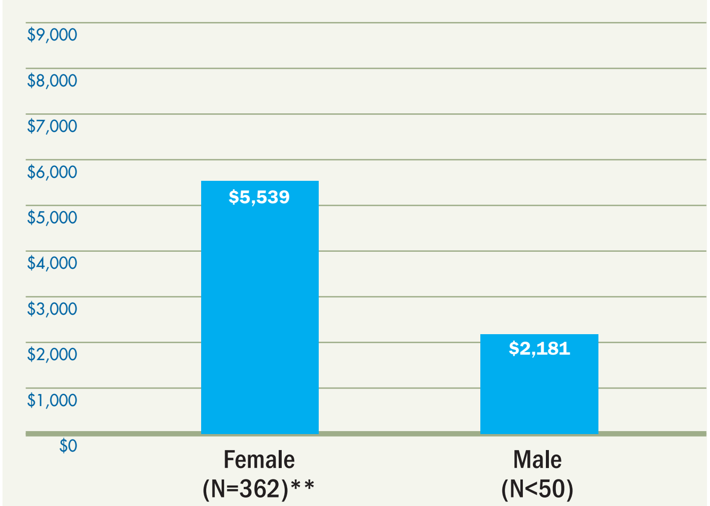
Nearly 80 percent of study participants were women, partly due to the fact that QUEST was targeting health care professions at the time. When we take a look at average earnings in the ninth year after enrollment, it's apparent that men were earning considerably more on average than women (Figure 5). Looking at this figure, one might reasonably conclude that if QUEST enrolled more men, the average salaries of participants would be higher.

Figure 5 Average Annual Earnings among Project QUEST Participants Nine Years after Random Assignment, by Gender



But when we examine the impact analysis (Figure 6), in which we compare QUEST participants to the control group, a far different picture emerges. It appears that QUEST's impact on female participants was much greater than it was on males, and if QUEST had enrolled only men, its overall impact on earnings would have been considerably smaller.

Figure 6 Project QUEST's Impact on Average Annual Earnings in Year Nine, by Gender

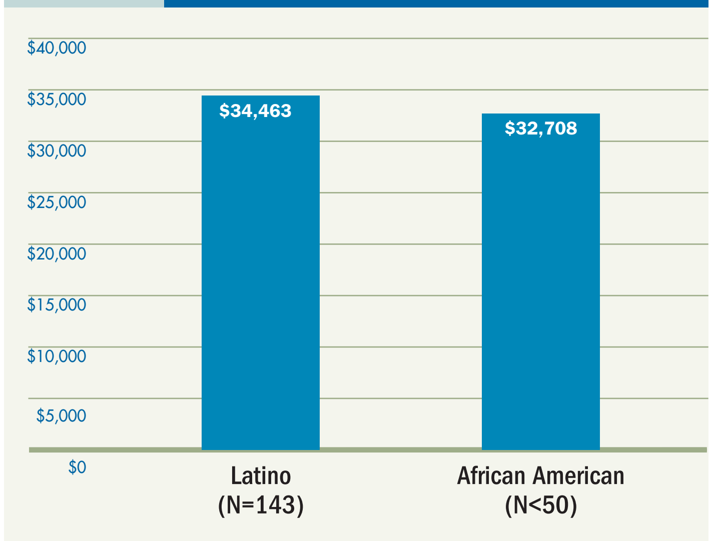


Note: Earnings impacts are reported in current dollars.
Within group significance levels: **p<.05

Race

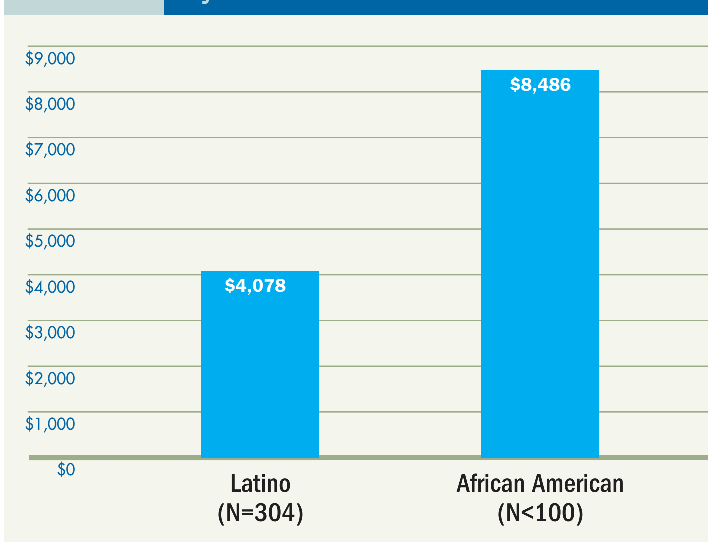
Nearly 90 percent of the study participants were Hispanic or African American. When we examine the ninth-year salaries of QUEST participants in these two groups, we can see that Hispanics made about $1,755 more than African American participants did (Figure 7).

Figure 7 Average Annual Earnings among Project QUEST Participants Nine Years after Random Assignment, by Race



But when we take a look at the impact analysis in Figure 8, it appears that African Americans benefitted more from QUEST's support than Hispanics did (though both groups still did quite well).

Figure 8 Project QUEST's Impact on Average Annual Earnings in Year Nine, by Race



As these examples illustrate, relying on outcomes is not sufficient for understanding how organizations can increase their impact—surely the most important measure of an organization's performance.[6] This is particularly important when the analysis and recommended strategies to improve performance focus on outcomes separated by demographic subgroups, such as age, gender, race, labor market experience, education, and criminal justice status.

Recommendations

For organizations

Understanding the difference between outcomes and impact is essential for all the staff of an organization, not just its senior leadership. Strategies to improve performance should be based on how to better serve all demographic groups, particularly those central to an organization's mission. Years ago, a well-known national technical assistance provider recommended that one of New York City's leading youth employment organizations stop serving people with GEDs, because its analysis of training completion rates indicated that GED-holders were less likely to graduate than participants who had a high school diploma. The organization's mission was to help disadvantaged youth, so it rightly questioned, and ultimately did not follow this advice.

Caitlyn Brazill underscores the challenge that organizations face in finding resources to serve people whose graduation and employment rates don't appear to be as high: "We have been seeking support to expand bridge programs, which enable applicants with literacy and math skills below the tenth grade [level] to take contextualized remedial instruction so they can enroll in our training. The challenge is that the cost per student is, of course, higher, but we may actually have a larger impact on those students than on those we do admit." Organizations should not abandon efforts to serve more challenging participants, and should make the case that philanthropic investments should be focused on impact.

For philanthropy

While it's certainly sensible to focus on outcomes, foundations should recognize the burden they collectively impose on organizations to collect and report their performance. Foundations should work together to identify the six to eight key indicators they all want from grantees. Any foundation that requires data collection beyond these common measures should pay for the extra costs of collection and analysis.

More important, foundations need to be wary of pushing organizations to improve performance simply by focusing on outcomes. While high performance goals are laudable, if they're achieved by excluding participants simply because they appear to achieve lower outcomes, organizations' ultimate impact may be diminished and their missions compromised.

Finally, when organizations such as Per Scholas, JVS Boston, and Project QUEST prove their impacts through rigorous evaluations, foundations should worry less about enforcing accountability measures and instead reward them by making substantial, less-restrictive investments to enable organizations to expand services, replicate in other locations, and pursue innovation in ways that respond to each organization's vision of its future.

Endnotes

1. https://www.bridgespan.org/insights/library/performance-measurement/building-capacity-to-measure-and-manage-performanc
2. These quotes are taken directly from foundation websites, but we did not feel it was important to identify them, as we view them simply as illustrative of a broader trend in philanthropy.
3. This is beginning to change, as recent rigorous evaluations of Per Scholas, JVS Boston, Project QUEST, Year Up, and Wisconsin Regional Training Partnership have demonstrated large, statistically significant earnings impacts.
4. https://economicmobilitycorp.org/wp-content/uploads/2019/04/NineYearGains_web.pdf
5. The subgroup analysis in the Project QUEST evaluation was exploratory. The study did not seek to enroll sufficient numbers of participants to be able to detect statistically significant differences within or between subgroups.
6. Judith Gueron shares similar findings comparing outcomes to impacts from the National Supported Work Demonstration in her 2016 paper "The Politics and Practice of Social Experiments: Seeds of a Revolution," pp., 9–10. https://www.mdrc.org/sites/default/files/2016_Gueron_MDRC_Working_Paper.pdf