Evaluation and Civil Society

LEHN M. BENJAMIN and DANA R.H. DOAN
Indiana University-Purdue University, Indianapolis USA

Key Words
- Evaluation
- Performance measurement
- Performance management
- Effectiveness
- Civil society
- Nonprofit
- Social impact
- Outcome measurement

Definition
Evaluating civil society organizations remains a pressing issue in the sector. Civil society leaders today face a dizzying array of models and approaches but evaluating the outcomes and impacts of these organizations remains challenging. Many of the changes civil society organizations (CSOs) pursue take a long time to address and are difficult to measure. Engaging in an evaluation requires making key decisions about the criteria, standards, and evidence that will be used to judge the merit or worth of the program, organization, or network. With multiple stakeholders, each with their own ideas about what may constitute successful change, determining who makes these decisions affects what gets measured, which affects how CSOs do their work. Despite these challenges, scholars and practitioners continue to work to develop more applicable and innovative approaches to assessing the contributions of civil society organizations. After giving a brief introduction and historical background to evaluation, this essay looks at some key issues in evaluating civil society organizations and concludes with recent developments on this important topic.

Introduction
Evaluation is a critical issue for civil society organizations (e.g., nongovernmental, nonprofit, private voluntary organizations). Policy makers, funders, activists, community members as well as civil society leaders themselves want to know whether these organizations are having an impact on the problems they seek to address. These concerns are not new and since the 1990s, we have seen an evolution of evaluative approaches in response to these concerns: from using logic models and theories of change to using social return on investment analysis and social impact bonds. What has remained consistent during this time is the demand for evidence about CSOs’ effectiveness, a demand that undoubtedly will continue as long as CSOs remain important partners for both governments and philanthropic donors in addressing social problems.

But evaluating the impact of civil society organizations is not easy. Many of the changes CSOs seek take a long time to address and are difficult to measure. At the same time, CSOs’ contribution to addressing these problems is partial and difficult to isolate given the variety of environmental factors that can also affect the desired outcome. For those CSOs who advocate for policy change, provide opportunities for civic engagement or organize communities, determining outcomes is even less tangible and more difficult to evaluate. Despite these challenges, those who study the field are working to develop more applicable and innovative approaches to assessing the contributions of civil society organizations to solving social
problems (e.g., Ebrahim 2019; Gugerty and Karlan 2018). After giving a brief introduction to evaluation, this essay looks at some key issues in evaluating civil society organizations and concludes with recent developments on this important topic.

**Historical Background**

Patton (1997) defines evaluation as the systematic collection of information to make judgments, improve effectiveness, and/or inform decision making about a program, initiative, or organization. The emphasis on the systematic collection of information sets evaluation apart from the informal evaluative judgments we make every day (Benjamin and Campbell 2020: 199). The emphasis on making judgments about the merit or worth of something requires making explicit the values guiding such judgements and is what distinguishes evaluation from basic research (Shadish, Cook and Leviton 1991). Improving effectiveness and informing decision making tells us that these judgments are made in service to a larger goal: to address some issue, improve performance, and to ensure everyone is able to thrive.

Although evaluation practice has been around a long time, the professional field of evaluation emerged in the 1960s during the War on Poverty (Scriven 1996:365). In those early years, evaluators focused on developing methodological approaches that could determine whether these new programs caused observable outcomes, using experimental and quasi-experimental designs, with the hope of informing future policy making (Weiss 1998). Over the years, those contributing to the field have developed theories pertaining to the core practice of evaluation: theories about knowledge construction, social interventions, values and valuing, as well theories about use and influence (Shadish, Cook and Leviton 1991). Today the field includes many approaches and models that address one or more of these theories. For example, theory-driven evaluation emerged to help stakeholders address the core assumptions underlying their program and its intended outcomes whereas utilization evaluation is premised on the idea that evaluation should be judged on its usefulness to intended users (Patton 1997).

Regardless of approach or epistemological orientation evaluation requires making certain key decisions: What criteria will be used to judge the merit or worth of the program, organization, or network (e.g., goal achievement, quality of implementation, satisfaction of beneficiaries, or financial strength)? What standards will be applied to these criteria (e.g., X percentage of those trained get full-time jobs, a majority of the residents believe the organization is effective in addressing blight)? And what evidence will be needed to understand whether such standards have been met (e.g., interviews or survey of trainees one or five years after they complete the program)? Of course, many other questions need to be addressed along the way, including: Who determines these criteria, standards, and evidence? Is there agreement among stakeholders over what constitutes success? If not, how is this resolved? Can sufficient evidence be gathered to understand the merit or worth of the program? Are the costs of collecting data feasible? What is the purpose of the evaluation and how will the results be used?

**Key Issues**

Early writings on evaluation and civil society focused on identifying the distinguishing characteristics of these organizations, compared to public agencies and for-profit firms. This work then considered the implications of these characteristics for evaluating CSO performance (e.g., Kanter & Summers 1987). Later researchers drew on organizational sociology to consider three approaches to assessing the effectiveness of CSOs: goal attainment, resource acquisition, and multiple constituencies views of the organization (e.g., Lecy et al. 2012). By the late 1990s and early 2000s researchers shifted their attention to how CSOs were responding to increasing demands for accountability and impact. At the time,
policymakers and funders were growing increasingly concerned about the effectiveness of CSOs, their limited capacity, and lack of accountability to communities. Consequently, they started to demand evidence of measurable outcomes (Edwards and Hulme 1996). Not surprisingly, researchers turned their attention to examining how CSOs were responding to these demands, including examining the prevalence of evaluation practices among CSOs and the challenges they faced in trying to measure the difference they were making in the lives of those they served (e.g. Benjamin 2008; Ebrahim 2003).

Given limited space, we organize this literature under three broad but inter-related questions1: 1) Are CSOs evaluating what they do? 2) What challenges do CSOs face as they try to evaluate their work? and 3) How is evaluation influencing CSOs? Before addressing these questions, it is worth distinguishing between evaluation and three closely related concepts: accountability, performance measurement, and effectiveness. In brief, accountability is one possible purpose for conducting an evaluation but by no means the only purpose. Other evaluation purposes include improving programs or processes and generating knowledge. Performance measurement involves the ongoing collection of evaluative data to inform organizational decision making and is one approach to evaluation. As noted above, evaluation includes a broad range of approaches. Effectiveness is one important criterion used to evaluate a program or an organization and is often defined as some level of goal achievement, but other criteria include efficiency, beneficiary satisfaction, responsiveness, and so on.

Are CSOs evaluating what they do?

The short answer is yes, if we take the results of several surveys in different parts of the world as an indicator. These surveys have, among other things, sought to understand the prevalence of different kinds of evaluative practices among CSOs, the types of data CSOs collect and who in the organization is responsible for collecting this data and carrying out the evaluation (e.g., Campos et al. 2011; Carman 2007; Hall et al. 2003; Marshall & Suarez 2014). All of these studies suggest that the majority of CSOs, greater than 50% of those surveyed, evaluate their work. Results also show that nonprofits use a variety of evaluation methods, including surveys, focus groups, interviews and administrative data. But the results also suggest a bifurcation in evaluation capacity among CSOs. For smaller CSOs, i.e., those not able to hire an evaluator, research suggests the quality of their evaluative efforts is low as they rely on program staff, with little time and limited expertise in evaluation, to design instruments and collect data or draw on existing administrative data (Hoefer 2000). Meanwhile, larger CSOs, with the ability to hire an internal or external evaluator, use sophisticated evaluation designs to assess impact (Campos et al. 2011).

However, using these studies as an indicator of whether CSOs are evaluating their work may be limited. Despite their usefulness, all the studies are one-time studies and employ unique survey instruments. Thus, we cannot compare across context or time. Moreover, the reported prevalence of evaluation practice among CSOs surveyed may be over-stated, as CSOs with limited capacity are less likely to complete such surveys. We also cannot be sure how evaluation terminology translates into different cultural contexts (Dahler-Larsen et al. 2017). All of these issues make it difficult at this time to have a firm grasp of the extent to which CSOs are evaluating their work. Such an assessment would require cross-country collaborations to develop simple instruments, that have common culturally appropriate questions, and that could be used consistently over time.

1 This summary is adapted from an expanded review. See Benjamin, Lehn M. & David A. Campbell. 2020. Evaluation and Performance Measurement. In Helmut Anheier and Stefan Toepler (Eds.), Routledge Companion to Nonprofit Management.
What challenges do CSOs face as they try to evaluate their work?

This question has received extensive attention by CSO scholars. This research suggests that CSOs face at least four challenges in evaluating their work. First, as noted above, CSOs have limited capacity to evaluate their work effectively. Consequently, the data they collect is low quality, often relying on internally developed survey instruments that have not been tested (e.g., Hoefer 2000). Second, the data they do collect is often in response to funders’ requests (e.g., Ebrahim 2003). Consequently, CSOs use the data primarily to report rather than improve their own practices. Moreover, the reporting relationship and the desire to show positive results to funders discourages collecting and using data in ways that would encourage organizational learning (Ebrahim 2003). And for CSOs reporting to multiple funders, all of which require distinct measures, this data collection and reporting can sap organizational time that might be used for other types of data-driven decision making (e.g., Benjamin, Voida and Bopp 2018).

Third, the recent interest by public and private funders in using experimental studies to evaluate CSO initiatives—whether through impact measurement, the use of evidence-based practice, or social impact bonds—has created isomorphic pressure for CSOs to follow suit, regardless of the kinds of programs and services provided, and often with disappointing results (see Silverstein & Maher 2008). Fourth, scholars have also suggested that existing program evaluation models—which assume an instrumental project logic—do not adequately reflect the emergent nature of outcomes, where outcomes are defined as nonprofits’ work in partnership with the communities they serve (e.g., Benjamin 2008).

How is evaluation influencing CSOs?

Mirroring some of the themes above, the research suggests that evaluation practice by civil society organizations has both intended and unintended consequences. The intended consequences mirror the two dominant evaluative purposes: using evaluation to inform practice and improve performance and using evaluation to demonstrate accountability to important stakeholders. As already noted, many CSOs use evaluation to demonstrate accountability, primarily to funders (Lee & Clerkin 2017). However, CSOs have also taken steps to ensure they are accountable to beneficiaries, including using feedback from beneficiaries to make actionable changes (see Twersky et al. 2013). Studies also find that CSOs use evaluation to inform practice and improve performance. For example, CSOs use data to make decisions about programs and services as well as to support organizational learning more broadly, particularly when they are able to link data collection and evaluation with strategic planning (e.g., Lee & Clerkin 2017). Still, reports suggest that CSOs underutilize the data they collect. One report noted that although 75% of nonprofits collect data, only six percent report that they feel they are using data effectively (Janus 2018).

In terms of unintended consequences, a number of studies identify such consequences, including inhibiting organizational learning, undermining participation by clients, underinvestment in relationship building with individuals and communities, to name a few. These consequences seem to be the result of the failure to fully account for how social change happens in these organizations and the dominant emphasis on using evaluation for accountability (e.g, Benjamin 2008; Ebrahim 2003). With that said, more recent developments, both inside and outside the evaluation field, offer promising approaches to address some of these concerns including: developmental evaluation, which assumes the identification of outcomes is emergent (Patton 2010); evaluation approaches that use systems theory to recognize complex and contingent change (e.g., Williams & Hummelbrunner
Future Directions

As the effort to evaluate the impact of CSOs continues, four recent developments, some of which we mentioned already, are worth noting. First, we see a growing emphasis on two types of evidence: evidence from randomized controlled trials (RCTs) and evidence that pairs social outcomes with financial and sometimes environmental outcomes. The trend towards RCTs is evident in large foundations and government aid agencies that fund international development. It is also evident in the adoption of evidence-based policy by governments that partner with human service organizations applying evidence-based practices proven with accumulated studies (e.g., Gugerty and Karlan 2018). Pairing social outcomes with financial outcomes is evident among private donors, particularly those coming from the corporate world, and some foundations. Interest in measuring impact through double and triple bottom line metrics, although not new, seems to appeal to these donors because they appear to offer a level of certainty and some control in achieving measurable impact. But both the elevation of randomized controlled trials and the coupling of social outcomes with financial outcomes has generated significant debate (Cooney & Lynch-Cerullo 2014).

Second, we see continued concern with issues of race, equity, and social justice in CSO evaluative efforts. Evaluation scholars have given significant attention to issues of equity and justice in evaluation (Donaldson & Picciotto 2016). Yet, in practice, evaluations can easily reinforce inequities by failing to involve all stakeholders, question dominant assumptions, or make value claims explicit (Sturges 2015; Thomas et al. 2018). In response a number of principles and approaches have been introduced to ensure that issues of power, privilege, and systemic inequity, are acknowledged and addressed in the process as well as the context of an evaluation (Dean-Coffey et al. 2014). Third, we see greater emphasis on collective and network-based approaches to evaluating impact, recognizing that nonprofits alone can rarely address community level problems. This requires evaluative approaches that mirror these efforts. Finally, with social media and advanced analytical technologies, the possibilities of big data to both increase effectiveness and advance understanding of effectiveness is getting attention in the CSO sector (Stiffman & Tumolillo 2019).

Cross-References

- Accountability
- Effectiveness and efficiency of nonprofit organizations
- Impact bonds
- Multiple bottom lines
- Performance and impact
- Social accounting for social economy organizations
- Social audits
- Stakeholders
- Values

References


Further Reading


of Voluntary and Nonprofit Organization, 26(5), 1805–1830.