Policy Briefs

Open Access

Katherine R. Cooper* and Michelle Shumate

Policy Brief: The Case for Using Robust Measures to Evaluate Nonprofit Organizations

DOI 10.1515/npf-2015-0029

Abstract: Although nonprofit organizations are expected to engage in continuous evaluation, its effectiveness is hampered by limited resources and competing and untested instruments. This paper makes the case for the creation and use of more robust measures in nonprofit evaluation. Specifically, we argue for the involvement of nonprofits in the development of reliable and valid instruments that can be used to benchmark nonprofit organizations against one another and for funders and government to support these efforts through their investment in nonprofit measurement. We cite a particular measure, The Nonprofit Capacities Instrument, as an exemplar.

Keywords: Nonprofit capacity, nonprofit measurement, evaluation

1 Introduction

Nonprofit organizations are encouraged or required to evaluate their activities or otherwise demonstrate their effectiveness, typically for purposes of indicating accountability to stakeholders (Benjamin 2008; Carman, 2010). However, nonprofit organizations struggle to evaluate their outcomes (Herman and Renz 1997); more recently, researchers have singled out social service organizations for their difficulties in measuring performance (Carnochan, Samples, Myers and Austin 2014).

The purpose of this paper is to present a case for improving the instruments available for nonprofits to evaluate their management practices and outcomes. This paper is organized as follows: first, the factors that hamper effective nonprofit evaluation are discussed, including limited resources to engage in evaluation, the

***Corresponding author: Katherine R. Cooper**, Department of Communication Studies, Northwestern University, 2240 Campus Drive, Evanston, Illinois 60208, USA, E-mail: katherine.cooper@northwestern.edu

Michelle Shumate, Department of Communication Studies, Northwestern University, Evanston, Illinois, USA

presence of competing instruments, and a lack of confidence in the measures themselves. These concerns are illustrated by drawing on the concept of nonprofit capacity and the creation of the Nonprofit Capacities Instrument (Authors 2015). Lessons learned from the development of the Nonprofit Capacities Instrument provide guidance for the development of nonprofit evaluation tools, including the involvement of nonprofits in the development and testing of instruments, the need for benchmarking, and the role of funders and governments in supporting the development of measures.

2 The Case for Measurement

Increasingly, nonprofit organizations are expected to report the outcomes of their efforts. There are several reasons for this; first, government agencies and funders expect – and may even rely upon – nonprofit organizations to reach populations that are underserved or marginalized and thus must account for their activities, which in turn informs evaluative activity (Benjamin 2008; Carman 2010). Additionally, nonprofit organizations are increasingly professionalized, suggesting that they face the same scrutiny as their for-profit counterparts and are expected to demonstrate more rigor in their own reporting (Hwang and Powell 2009). Hwang and Powell also suggest that nonprofits are increasingly competing with other agencies for resources; evaluation results may also be involved in allocating these resources.

Much of the work on nonprofit measurement can be found within the evaluation literature. This body of work suggests that nonprofit measurement is commonly addressed, but that questions remain as to which instruments and evaluative activity are helpful to nonprofits. For example, Snibbe (2006) suggested that although evaluative activity is increasing, organizations are overwhelmed with data that is not necessarily useful to them. Thomson's (2010) review of previous studies suggested that there is a gap between what nonprofits intend to evaluate and what they actually do, and that there is "significant room for growth" (p. 4). Thomson's own (2010) study suggests funder mandates may increase outcome measurement in nonprofits, but ultimately concluded that the more important question whether organizations use this information in decision-making. The interpretation and use of data continues to be problematic for those collecting mandating, funding, or collecting data; Liket, Rey-Garcia, and Maas (2014) describe an example in which the funder was mistakenly under the impression that evaluation data captured program effectiveness and was used for program improvement, when in fact the evaluation results had never been used for that purpose.

These studies suggest that nonprofit stakeholders differ in their understanding of why something is measured. Behn (2003) refers to several possible

purposes of evaluation, some of which might be useful to foundations (e.g., evaluation as budgeting in order to allocate resources), government (e.g., evaluation for compliance), clients and individual donors (e.g., evaluation as promotion to demonstrate the organization's success to stakeholders), and to the organization itself (e.g., evaluation for the purpose of learning and improving). These different approaches suggest a range of reasons why nonprofits conduct evaluation but these distinctions may not be clearly communicated across stakeholders. If these interests are not aligned with one another, resources dedicated to completing nonprofit measurement are ultimately wasted.

3 Challenges in Nonprofit Measurement and Evaluation

In addition, there are several problems with existing measures that make it difficult to report on nonprofit operations or outcomes. First, limited nonprofit resources often make it difficult to use or interpret existing measures. Carman and Fredericks (2010) suggested that nonprofits' ability to carry out evaluation activity varies widely. Their analysis suggested three groups of nonprofits: those that successfully engage in evaluation, those that struggle to engage in evaluation activities beyond funder requirements, and a third group that struggles to engage in any evaluation at all. The results suggest that nonprofits are challenged by a lack of technical capacity for conducting evaluation (Carman and Fredericks 2010); additionally, many nonprofit organizations do not have the time or the financial resources to undergo formal evaluation procedures. For example, Liket, Rey-Garcia and Maas (2014) note that nonprofit evaluation is often undertaken by a program manager or other staff member because there is no available budget for external evaluation. This is especially problematic when evaluation tools require a trained facilitator or an extensive time investment to complete. Moreover, instruments that require extensive interpretation by trained evaluators are unusable by nonprofits that lack the resources to hire such an expert, especially for repeated evaluation of capacity building or quality improvement efforts.

Second, many competing measures exist, using different different definitions and models of the object of evaluation. Many of these can be found in catalogs or archives (e.g., The Foundation Center n.d.(a,b); PerformWell n.d.). Such sites offer extensive lists of resources; however, the presence of so many instruments renders it difficult for a nonprofit to select the appropriate instrument. This task is further complicated by the fact that many of the existing instruments are similar to one another. For example, the Foundation Center

Issue Lab portal includes over 300 resources related to evaluating board effectiveness. Most are embedded within agency or consultant reports, which makes the process of identifying the right tool for board evaluation more arduous.

Third, many measures available to the nonprofit community, and mandated by funders, have no evidence of reliability or validity. Reliability means that the items in a measure or the whole measure over time consistently produces the same result. Validity refers to evidence that the instrument actually measures what it claims. There is a movement to better capture information on how measures have been developed and applied; for example, PerformWell (n.d.) offers background on how various measures were developed and tested and allows users to rate the tools. However, the measures included do not evaluate nonprofit management or systems; Information on the reliability or validity of most nonprofit management or operations instruments, developed by consultants, foundations, or nonprofits and in wide use, are not readily available. As such, no funder or regulator can be confident that these measures are objective.

4 The Development of the Nonprofit Capacities Instrument

In developing the Nonprofit Capacities Instrument, we encountered all three of these challenges. Many existing evaluation measures of nonprofit capacity required an external evaluator and several days to complete. There were dozens of measures, each with its own definition and model of nonprofit capacity. And, none of the instruments had evidence of reliability or validity. To address these deficiencies in measurement, we invested in the development of a self-administered, quantitative measure of nonprofit capacities.

To develop the measure, we followed guidelines suggested by Worthington and Whittaker (2006) and DeVellis (2012), using an inductive-confirmatory two-study approach. First, we conducted a thorough review of the literature that enabled the research team to develop a definition of the concept being measured, and created an item pool comprised of items compiled from existing capacity instruments (see Authors 2015). We tested this pool of items by surveying small to mid-size nonprofits in two geographically bounded areas, one in United States and one in Costa Rica. This required translation and back translation of the measure from English to Spanish. We used exploratory factor analysis and inter-item reliability analysis to yield a refined instrument. We also gathered data using the SOCAT, a qualitative measure of capacity (Grootaert and Van Bastelaer 2002) to determine if any additional information emerged from the

qualitative data that was not captured in our item pool. In this same sample, we collected data on peer-ratings and self-reported ratings of nonprofit effectiveness for purposes of determining if the measure had criterion validity, or to determine whether the scale in question exhibits an empirical relationship with a standard measure (DeVellis 2012).

The refined instrument was disseminated in four languages, English, Spanish, French and simplified Chinese, to a second international sample, this time of mid-size to large nonprofits. We used confirmatory factor analysis to demonstrate the criterion validity of each of the sub-scales (i.e., that the eight subscales were measuring different things). Additionally, by replicating the results in a new sample, as suggested by Worthington and Whittaker (2006), we were able to have even greater confidence in the measure. The detailed version of the Nonprofit Capacities Instrument's development is chronicled more thoroughly by Authors (2015); we describe the process here to illustrate some of the challenges – and possibilities – in developing robust measures for nonprofit evaluation.

5 Lessons in the Development of Nonprofit Evaluation Instruments

Drawing from our experience in developing the Nonprofit Capacities Instrument, we suggest three implications for those seeking to fund or evaluate nonprofit projects. First, nonprofit organizations must be involved in the creation of instruments, and instruments should be developed for a nonprofit context. Thomson's (2010) research suggests that there are gaps between what nonprofits aim to do and what they can actually accomplish in evaluation, and that we know less about whether nonprofits can use these input in their decision-making. Research also suggests that tools for performance management are often developed in the public or corporate sector and then appropriated for nonprofit use (Carnochan et al. 2014; Ospina, Diaz and O'Sullivan, 2002). However, if nonprofit evaluation tools are created explicitly for nonprofits – and they are involved in their development and testing – we have a better chance of developing instruments that are actually useful to the nonprofit. In our work developing the Nonprofit Capacities Instrument, we took several steps to ensure that the measure was nonprofit-specific and that nonprofits were involved in developing and testing the measure. To make sure that we were drawing from nonprofit instruments, we created an extensive item pool from existing measures of nonprofit capacity. To ensure that nonprofits could be involved in

testing, we conducted two waves of instrument validation across a diverse sample. The instrument was translated into four languages and tested across different categories of organizations and an international sample to ensure that the measure was consistent across these variations. Additionally, this project had a nonprofit advisory team who shared their concerns and offered feedback during the development of the instrument.

Second, nonprofit organizations benefit from the use of benchmark data, and robust measures should enable organizations to compare their findings to others. Benchmarking refers to a “systematic, continuous process of measuring and comparing an organization’s business processes against leaders in any industry to gain insights that will help the organization take action to improve its performance” (Saul 2004, p. 7). Benchmarking has been described as a tool for promoting organizational learning within nonprofits (Buckmaster 1999) and is described as a strategy for nonprofit management. Numerous resources for benchmarking exist, including Saul (2004) and Keehley and Abercrombie (2008). However, despite its proclaimed usefulness, it is uncertain as to the extent to which nonprofits engage in this activity; Conley Tyler (2005) found numerous challenges to benchmarking across nonprofits in Australia and concluded that the lack of benchmarking was similar to other countries. We suggest that robust measures that are reliable and valid should also be widely applicable to further aid nonprofits in benchmarking, which benefits individual organizations as they work towards its goals and also provides an assessment of the field that may be useful to funders, evaluators and practitioners. Unlike previous capacity instruments that reported individualized measures, the Nonprofit Capacities Instrument provides benchmarks that indicate an organization’s capacity in comparison to others and was developed for this purpose. For example, each organization that completed a report received an assessment that compared their findings to other organizations within their service area, geographic region, and size as measured by organizational assets. Reliable, valid instruments provide a sense of where a nonprofit organization is in relation to others and may ultimately be more helpful than individualized, qualitative assessments that do not capture the broader environment in which the nonprofit works.

Third, the development of robust measurement within the nonprofit sector requires investment from governments or foundations. Previous studies have examined the influence of government (Carman and Fredericks 2008) and funders (Thomson 2010) in requiring and funding evaluation. However, in order to improve measures available to nonprofit organizations, funders and government must go beyond simply paying for or encouraging nonprofit measurement; rather, these agencies should encourage the development and testing

of reliable and valid instruments. The guidelines previously suggested – the involvement of the nonprofit sector in creating measures and the inclusion of benchmark data – are possible only if external stakeholders see their value. Our examination of nonprofit capacity indicated that foundations and international agencies often created their own resources for building and evaluating capacity (see Foundation Center n.d.; USAID Center for Development Information and Evaluation 2000), but failed to take the next step in establishing these measures reliability and validity; ultimately, Authors' (2015) development of the Nonprofit Capacities Instrument was made possible by support from the National Science Foundation. This funding enabled us to build a research team with the capacity to conduct extensive reviews of the literature and existing measures, recruit an international sample, perform multiple rounds of rigorous empirical testing, and provide detailed assessments to each nonprofit that participated in the instrument. Additionally, funding enabled us to compensate the nonprofit organizations that were involved in the time-consuming process of testing what was, in the first wave of the study, a lengthy instrument.

This is an expensive and admittedly arduous process. The Nonprofit Capacities Instrument represents a five-year investment in the literature and empirical approaches to nonprofit capacity. However, evaluators may consider this a worthwhile investment that builds relationships between nonprofit funders and practitioners, and ultimately builds confidence in the results of the assessment.

6 Conclusion

Although nonprofit organizations are encouraged to participate in evaluation, research suggests that this may be more of a formality than a useful exercise. We suggest that part of the problem in nonprofit evaluation is the measures themselves; specifically, limited nonprofit resources for evaluation, the presence of similar, competing instruments, and a lack of empirical development that results in few reliable and valid instruments.

We suggest three directions for the development of nonprofit measures, including the involvement of nonprofits in the creation of these instruments and the inclusion of benchmarking as a strategy to improve upon measurement and evaluation within individual organizations and the nonprofit sector as a whole. These resources are possible with foundation and government investment.

Throughout this study, we refer to the development of the Nonprofit Capacities Instrument, which demonstrates the challenges in instrumentation

as well as the possibilities for nonprofit measurement. Developing the Nonprofit Capacities Instrument was a costly, time-consuming endeavor, but the project resulted in the first reliable, valid measure of nonprofit capacity. As a result, nonprofit organizations, as well as the clients they serve and funders who support them can be confident that the instrument truly measures nonprofit capacity, and that this measure can be used across organizations regardless of size, location, or mission. It also fills a need for nonprofit leaders, who can now use this instrument to evaluate and re-evaluate their capacity over time.

Although consultants and corporations provide tools that may be helpful to nonprofit work, reliable and valid instruments provide more accurate and dependable assessments. Such instruments demystify nonprofit activities and elevate the nonprofit sector as a whole; however, the development of such instruments is possible only with the support and commitment of foundations and government.

References

- Shumate, Michelle, Katherine, R. Cooper, Andrew Pilny, and Macarena Peña y Lillo. 2015. "The Nonprofit Capacities Instrument." Paper presented at the annual meeting of the Academy of Management, Vancouver, Canada, August 7–11, 2015.
- Behn, Robert D. 2003. "Why Measure Performance? Different Purposes Require Different Measures." *Public Administration Review* 63:586–606.
- Benjamin, Lehn M. 2008. "Account Space: How Accountability Requirements Shape Nonprofit Practice." *Nonprofit and Voluntary Sector Quarterly* 37 (6):201–23.
- Buckmaster, Natalie. 1999. "Benchmarking as a Learning Tool in Voluntary Non-Profit Organizations: An Exploratory Study." *Public Management an International Journal of Research and Theory* 1 (4):603–16.
- Carman, Joanne G. 2010. "The Accountability Movement: What's Wrong with This Theory of Change?" *Nonprofit and Voluntary Sector Quarterly* 39 (2):256–74.
- Carman, Joanne G., and Kimberly A. Fredericks. 2010. "Evaluation Capacity and Nonprofit Organizations Is the Glass Half-Empty or Half-Full?" *American Journal of Evaluation* 31 (1):84–104.
- Carman, Joanne G., and Kimberly A. Fredericks. 2008. "Nonprofits and Evaluation: Empirical Evidence From the Field." *New Directions for Evaluation* 119:51–71.
- Carnochan, Sarah, Mark Samples, Michael Myers, and Michael J. Austin. 2014. "Performance Measurement Challenges in Nonprofit Human Service Organizations." *Nonprofit and Voluntary Sector Quarterly* 43 (6):1014–32.
- Conley Tyler, Melissa. 2005. Benchmarking in the Non-Profit Sector in Australia." *Benchmarking: An International Journal* 12 (3):219–35.
- DeVellis, Robert F. 2012. *Scale Development: Theory and Applications*, Vol. 26. Thousand Oaks, CA: Sage publications.
- Foundation Center. n.d.(a) "Capacity Building for Nonprofit Organizations: A Resource list." Retrieved from <http://foundationcenter.org/getstarted/topical/capacity.html>

- Foundation Center. n.d.(b) Topical “Resource Lists: Evaluation for Nonprofits.” Retrieved from <http://foundationcenter.org/getstarted/topical/eval.html>
- Grootaert, Christiaan, and Thierry Van Bastelaer, eds. 2002. *Understanding and Measuring Social Capital: A Multidisciplinary Tool for Practitioners*, Vol. 1. Washington, DC: World Bank Publications. 2002.
- Herman, Robert D., and David O. Renz. 1997. “Multiple Constituencies and the Social Construction of Nonprofit Organization Effectiveness.” *Nonprofit and Voluntary Sector Quarterly* 26 (2):185–206.
- Hwang, Hokyu, and Walter W. Powell. 2009. “The Rationalization of Charity: The Influences of Professionalism in the Nonprofit Sector.” *Administrative Science Quarterly* 54 (2):268–98.
- Keehley, Patricia, and Neil Abercrombie. 2008. *Benchmarking in the Public and Nonprofit Sectors: Best Practices for Achieving Performance Breakthroughs*. San Francisco: John Wiley & Sons.
- Liket, Kellie C., Marta Rey-Garcia, and Karen E.H. Maas. 2014. “Why Aren’t Evaluations Working and What to Do About It: A Framework for Negotiating Meaningful Evaluation in Nonprofits.” *American Journal of Evaluation* 35 (2):171–88.
- Ospina, Sonia, William Diaz, and James F. O’Sullivan. 2002. “Negotiating Accountability: Managerial Lessons from Identity-Based Nonprofit Organizations.” *Nonprofit and Voluntary Sector Quarterly* 31:5–31.
- PerformWell (n.d). Retrieved from <http://performwell.org/>
- Saul, Jason. 2004. *Benchmarking for Nonprofits: How to Measure, Manage, and Improve Performance*. St. Paul, MN: Fieldstone Alliance.
- Snibbe, Alana Conner. 2006. “Drowning in Data.” *Stanford Social Innovation Review*: 39–45. Retrieved from http://www.ssireview.org/articles/entry/drowning_in_data
- Thomson, Dale E. 2010. “Exploring the Role of Funders’ Performance Reporting Mandates in Nonprofit Performance Measurement.” *Nonprofit and Voluntary Sector Quarterly* 39:611–29.
- USAID Center for Development Information and Evaluation. 2000. Measuring institutional capacity. *TIPS*, 15, http://pdf.usaid.gov/pdf_docs/PNACG612.pdf.
- Worthington, Roger L., and Tiffany A. Whittaker. 2006. “Scale Development Research a Content Analysis and Recommendations for Best Practices.” *The Counseling Psychologist* 34:806–38.