Introduction

A recent article in the Stanford Social Innovation Review observed that, in the name of accountability, funders, donors, and the general public are calling for more program evaluations by not-for-profit agencies. And it argued that, as a result of this “… nonprofits often are collecting heaps of dubious data at great cost to themselves and ultimately to the people they serve.”

This paper makes two points and elaborates on the second. First, as the article notes, the question of when and how to evaluate programs is an extremely complex one, and one that evaluators have done little to help funders, nonprofit executives and practitioners understand. Second, nonprofits provide funders reams of data concerning their finances, activities, and products in formats and for purposes that often are idiosyncratic to each funder, arbitrary, and virtually impossible to explain or justify – but certainly have nothing to do with evaluation or, for that matter, with managing programs and helping people. Thus, the outgoing president of one of this country’s largest foundations once described with a sigh, having discovered that a program officer had taken a grantee to task for spending grant funds on a single pack of pencils! Forcing nonprofits to over-measure what they do, that is, to collect reams of accounting, process and/or activity data is prevalent, foolish and wasteful. More importantly, over-measurement erodes their ability to focus on mission-critical work.

Between a Rock and a Hard Place: Forming Destructive Habits
Why all this data collection? No doubt for many reasons. Surely right at the head of any list is the strong desire to be accountable…and little understanding of how to go about it. This leads funders and nonprofit agencies to adopt two very destructive habits: know-nothing optimism on the one hand, and mind-numbing over-measurement on the other.

Know-Nothing Optimism

Consider the following mission statements of some leading nonprofit organizations taken recently from their websites:

- **The mission of YouthBuild USA is to unleash the intelligence and positive energy of low-income youth to rebuild their communities and their lives.**

- **Camp Fire USA builds caring, confident youth and future leaders.**

Well...do they deliver? The truth is, since neither has been evaluated in a rigorous manner, nobody really knows. Of course they are not alone. Very few organizations in the social sector (funders and not-for-profit organizations) actually try to evaluate the results of their efforts. So know-nothing optimism – over-claiming and under-measuring – has become the sector norm. And surely this stems, in no small part, from the justified sense that, as organizations enjoying preferred public policy status, they really do need to justify their existence. But if organizations make such optimistic claims, and because of favorable tax exemptions use revenues that otherwise would be available to government to undertake such work, don’t they have an obligation to assess and communicate what they actually accomplish rather than merely describing what they do or hope to achieve? Surely, society deserves – needs – a rational account of what the social sector really is accomplishing.

However, know-nothing optimism isn’t the whole picture. In fact, government, some private funders, and even a growing number of not-for-profit agencies are increasingly recognizing the compelling public policy rationale for evaluation within the social sector. Sometimes the call to arms is to the tune of “accountability”, other times it is “effectiveness.” But more often than not both induce dysfunctional behavior.

---

Here it is a story that is about both issues: tackling the question of program effectiveness and also assuring public accountability. In 2003, Mathematica Policy Research, Inc. (one of the America’s most prominent evaluation organizations) released its first year report of the 3-year evaluation of the US Government’s 21st Century Community Learning Centers after-school program, which was designed to enhance academic skills and achievement for (mostly poor) middle school children, and is offered at some 1,500 sites nationally. This report clearly stated its findings were very preliminary and should not be used for policy-making purposes; but it did indicate a preliminary finding that there was no evidence of improved academic outcomes for participants.

Immediately upon its release, the George W. Bush administration announced plans to cut federal funding for this program by $400 million – a full 40% of the $1 billion that had been dedicated to it. In the resulting uproar several troubling things happened: (a) the Mathematica study was subjected to intense criticism by leading youth development advocates; (b) these advocates, in lobbying to have the funding restored, adopted a dismissive stance toward the use of evaluation data for setting funding priorities; (c) members of the evaluation profession were also drawn into the fray of schism, with some outspokenly attacking (using uncharacteristically dogmatic and harsh language) the methods and validity of the Mathematica study, and others rising to Mathematica’s defense; and (d) the funding ultimately was restored as a result of intensively strident and emotional lobbying by “youth advocates” – in which the legitimate uses of evaluation data for informing social policy development and funding decisions were left undiscussed, and consequently once again relegated to the margins of social utility.

In short, a disaster! The forces of know-nothing optimism prevailed over the equally dismal forces of knee-jerk high-stakes accountability measures. So...what is the social sector’s answer to know-nothing optimism? Unfortunately, quite often it is the equally dysfunctional practice of mind-numbing over-measurement.


4 In fact, subsequently the final Mathematica Report confirmed the finding that academic achievement was not a statistically significant impact of this program, which actually seemed to harm children a bit by increasing the likelihood that they would express frustration and anger by breaking things. See Dynarski M., James-Burdumy S., *et al.*, “When Schools Stay Open Late: The National Evaluation of the 21st Century Community Learning Center Program,” Washington D.C: U.S Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance (2004)
Mind-numbing Over-measurement

Here we encounter a paradox: On the one hand, very few funders or the grantees they support have worked out clear and measurable objectives and indicators for assessing success. On the other hand, as was noted above, not-for-profit agencies and their funders are earnestly collecting and reporting on mountains of data in the Sisyphus-like pursuit of...what?

Consider the following examples:

**Turnstile numbers.** Many organizations – with the encouragement of their funders – commit significant staff time and resources to count each and every person with whom they have any kind of contact...or whom they have “touched.” Some even count website hits and then proudly and publicly report this data.

**Outcomes.** Over the past several decades, pressured by both the Federal Government (through the “Government Performance and Results Act” of 1993) and the efforts of the United Way, public and social sector agencies have been working frantically and at considerable cost to collect outcome data. These data, too, are reported proudly and publicly.

But to what end?

Turnstile numbers are little more than a way of suggesting how much work an organization may be doing. But even for this limited purpose, turnstile numbers are weak – after all, “touches” via a website or registering people for programs they might attend once or twice or not at all will never add up to much that is of value. So turnstile numbers alone *most certainly* don’t tell the story of what an organization really is accomplishing.\(^5\) Lots of counting, little shedding of light.

---

How about outcomes? They do tell more of a story than turnstile numbers…but let the listener beware. A thought-provoking article describes the outcome data for programs seeking to help single mothers get off welfare by preparing them for employment and getting them jobs. The outcome data for the program in Grand Rapids, Michigan, showed 76% of its participants working. The outcome data for Atlanta, Georgia, were 57%; and for Riverside, California, 46%. Looks good for the Grand Rapids program – until their outcomes are compared with the outcomes for a randomly selected “control group” from the same city…which showed 61% working. In Atlanta, 53% were working; and in Riverside, 38%. So: which program had the greatest impact? Which created the most social good? Clearly the answer is Riverside, which improved the employment rate of its participants 8% (versus 6% for Grand Rapids and 5% for Atlanta). In other words, collecting and reporting outcome data alone, although a triumphantly quantitative exercise, can lead to spurious precision when it is used to assert the effectiveness of a program and to promote its claims of success - more counting, much misleading.

So…what is to be done? Is there a way to navigate between the serpentine Scylla of know-nothing optimism and the engulfing Charybdis of mind-numbing over-measurement? In a nutshell, the answer is for nonprofit organizations and funders to build and use strong internal performance management systems to learn from and improve the quality of their work.

---


7 Here I am not arguing for subjecting all programs to randomized control trials to prove their effectiveness. Far from it. Such costly and time-consuming experiments should only be undertaken when conditions are right and there is a clear strategic reason for doing so. See Hunter D.E.K. “Daniel and the Rhinoceros,” Evaluation and Program Planning (May 2006): 180-185
Performance Management: The KISS Approach
(Keep It Simple, Stupid)
Bring up performance management with social sector leaders and you’ve got a good chance of watching their eyes glaze over. It conjures up the worst, dehumanizing practices of the corporate sector, and reeks of rampant data gathering run amok...both of which represent the sector’s worst fears.

But what is performance management? Simply put, **performance is intentional action dedicated to reaching one or more measurable objectives**. Performance can be good or bad, effective or ineffective, successful or unsuccessful, well-managed or not. **Performance management, then, is a multi-step, self-correcting process** that focuses on driving an organization with intentionality toward explicit, clear, measurable objectives, instituting ways to learn from front-line and managerial experience along the way, clarifying what it is doing and achieving, adjusting activities as needed, and even revising the original objectives if indicated.

Here is where some solid, useful advice can make all the difference – starting with the military acronym **KISS**: *Keep It Simple, Stupid*. If an organization really knows what it is doing and why – if it has a theory of change that is meaningful (to key stakeholders), plausible (in that it makes sense to both stakeholders and content area experts), doable (within the resources and capacities of the organization and, perhaps, its strategic partners), and assessable (that is, with measurable indicators of progress and success) – then designing simple, useful, performance metrics really isn’t forebodingly hard.\(^8\)

This will not have escaped anybody who has been treated in a hospital within the last decade. Inevitably, patients will be asked: “On a scale of 1 to 10, how much pain are you in?” Pain is, of course, totally subjective and infinitely variable to each person’s sensitivity to and threshold for managing it (not to mention cross-cultural variation in the meaning and experience of pain)...yet clinicians have learned that following each patient’s movement along the scale of experienced pain leads to much improved medical responsiveness and therefore improves the quality of treatment – a worthwhile objective grounded in a straightforward theory of change (the medical understanding of pain’s relationship to physical pathology) and a performance metric of “KISSable” simplicity!

Once practitioners are freed from their aversion to numbers they can use their knowledge, experience, and skills to develop a few indicators to track in order to assess whether their programs are making progress and succeeding in achieving their objectives. Take the example of Roca Inc. (an organization serving mostly immigrant,

---

street life and gang-involved young people in and around Chelsea, Massachusetts). Roca believes that, to help these young people, it must engage them in five to six year-long “transformational relationships” with caring staff who work relentlessly to help them succeed in education and work. To help their youth workers keep focused on what they need to do, Roca not only measures and manages progress toward these outcomes, but also the progress of each “transformational relationship” – up the ladder from “pre-contemplation” to “contemplation,” then “planning,” “action,” and finally, “sustaining”. These ratings are used for purposes of supervision and as an ongoing assessment of the quality and effectiveness of how each youth worker is working with each and every young person. Simple? Yes. Simplistic? No! Useful (for Roca's purposes)? Very!! And that is what performance management is all about: identifying and implementing means to assess, track, and use what is necessary to improve the quality and effectiveness of organizations, programs and the work of individual front-line staff.

*Performance management is at least one way to navigate between over-measurement and unsubstantiated optimism, to grapple with what is real and what is illusory and to take seriously and respond meaningfully to just demands for accountability.* It also provides a framework for funders and not-for-profit organizations to agree on simple and useful metrics that can inform evaluative learning and point to the social consequences of grantees’ efforts and funders’ investments (grants).
Steps to Successful Performance Management

Contrary to what many believe, implementing a performance management system is not simply a matter of building or choosing a database, a subject that will be addressed shortly. Rather, it involves a set of critical, interdependent actions on the part of an organization. Here they are presented in sequence, but of course they are highly interactive:

**Step 1**

The organization’s leadership (executive team and board) embrace the need to manage the organization rationally, efficiently and with ongoing attention to the quality and effectiveness of its work. (This is not a foregone conclusion in the social sector.)

**Step 2**

The organization develops an organizational Theory of Change (ToC) regarding how it creates and sustains the delivery of socially valuable goods and services. The ToC articulates (a) the domain within which the organization works, (b) the results it is working to achieve and the indicators it will use to assess its outcomes, and (c) the program and/or services it will provide to achieve its outcome objectives. Such a ToC addresses far more than program logic models; rather, it looks at the organization as a whole, inventories its capacities and developmental needs, and itemizes its resource requirements so that it can create its goods and deliver its services reliably, at high levels of quality, and over the long haul.

---


ii. See reference in footnote 8.
Steps to Successful Performance Management

Step 3
Using the ToC as a framework, the organization develops key metrics and the targets it will use to track its performance, both organizationally and within programs, and builds processes for doing so and using this information to make ongoing adjustments as indicated. Part of this work involves bringing staff at all levels of the organization to a point where they understand the outcomes, indicators, and performance targets well, and are committed to working in ways that drive up their personal performance and the performance of the organization as a whole.

Step 4
The organization implements a performance management data system, and uses it to manage work at the individual, program, and organizational levels – in real time – to achieve the highest degree of success possible. This requires some key fundamental decisions: (a) to build or to buy a software application, and (b) to run it in-house or to outsource it. The appendix presents key elements that an organization should consider in choosing a performance management data system, and the following words of wisdom seem worth quoting in this regard: “Any software that is being considered must balance the day-to-day needs of direct service staff and their supervisors who play a fundamental role in service delivery and a program’s performance, and the strategic reporting needs of the organization. The tool should be easy for them to use, relevant to both their jobs and their experiences, and make the difficult job they are asked to do easier by allowing them to better organize and manage information. Data for researchers and evaluators should be a by-product of the tool, not the reason for its existence.”

Step 5
The organization routinizes use of performance data to learn from and deepen the organization’s understanding of its work, to align the staff and leadership of the organization behind shared performance expectations, to learn from experience and make data-driven adjustments as indicated, and to invigorate the organization’s commitment to working relentlessly, passionately, and productively to achieve its mission.

iii. Personal communication from Steve Butz, Founder and President of Social Solutions, the company that markets Efforts to Outcomes, one of the cutting edge, web-based performance management data systems.

iv. And, when circumstances are right – when the programs and services an organization delivers are implemented and managed to high levels of performance, have been done so reliably and for years on end, and when there is a strategically compelling reason to do so – the organization might consider the utility of undertaking a formal evaluation of its impacts.
A critical point: **The worst error to make in implementing the five steps outlined here is to do so in a top-down, command and control manner...and to implement a performance management data system that mirrors this.** Rather, organizations that have been successful in using performance management have recognized how profoundly transformative it is...that it is really a giant step forward in an organization’s development and that the organization can never again be the one it was before undertaking this process. In recognizing this, the leaders of such organizations take the time and dedicate the resources necessary to include staff at all levels and within all key roles of the organization in the processes outlined above.

Experience shows that most organizations will need three to six years to do so. A considerable investment...but one that the organization’s intended beneficiaries fully deserve.

**Evaluation versus Performance Management**

Before ending this paper it seems worthwhile to address the fact that performance management often is confused with evaluation. While they are certainly interrelated, there are real differences between the two. The following table captures some of the key differences in the nature of the data each uses.  

<table>
<thead>
<tr>
<th>Performance Management Data</th>
<th>Versus</th>
<th>Evaluation Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program Planning, Implementation, and Management</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Evolving/Iterative</td>
<td>Purpose</td>
<td>Negotiated up front</td>
</tr>
<tr>
<td>Broad: tracks all operations</td>
<td>Focus</td>
<td>Issue-specific</td>
</tr>
<tr>
<td>Constant/ongoing</td>
<td>Frequency</td>
<td>Periodic</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Performance Management Data</th>
<th>Versus</th>
<th>Evaluation Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process and Content: Inputs, Outputs, Outcomes, Indicators, Measures</td>
<td>Types</td>
<td>Content: Inputs, Outputs, Outcomes, Impacts</td>
</tr>
</tbody>
</table>

### Obtaining Data

<table>
<thead>
<tr>
<th>“Real time” or ongoing</th>
<th>Data Production</th>
<th>Rare – at large period intervals (monthly/quarterly/yearly)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Presumed</td>
<td>Data Validity</td>
<td>Must be established as part of the research</td>
</tr>
</tbody>
</table>

| Internal data (demographic, staff efforts, client service utilization, outcomes), assessments, surveys, etc. | Tools | Desk research, interviews, surveys, information systems, etc. |

### Using Data

<table>
<thead>
<tr>
<th>Presumed for management purposes</th>
<th>Establishing attribution of causality</th>
<th>The core issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Real time” or ongoing</td>
<td>Development of Organizational learning</td>
<td>Infrequent</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>High (&quot;real time&quot; flow of data)</th>
<th>Tactical Utility</th>
<th>Low (very slow flow of data)</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Low</th>
<th>Strategic Utility</th>
<th>High</th>
</tr>
</thead>
</table>
This paper is about the kinds of data described in the left column. Any nonprofit agency can develop such data easily once it has clarified its strategy and theory of change – that is, once it really knows *why* it is in business and *how* it intends to work to achieve its goals and objectives, with indicators for outcomes and perhaps a few performance standards as the key metrics. They need not become research organizations to do this, nor even hire evaluators. And they need not drown in data if they keep data gathering simple and focused on essentials. But they do need to be convinced that, only by measuring and using key metrics to manage their work, can they be sure that their efforts will yield the intended benefits to the people they serve.

To do less hardly seems ethical.
Appendix: Performance Management Data Systems – Key Elements to Assess

- **Cost** – software package price or initial license fee, implementation charges, yearly license fee, charge per end user, etc.

- **User friendliness** – ease with which interface can be learned, intuitive navigation/menu structure, degree that screens look clean and inviting to use, ease with which each end user can customize/configure screens within the software without paying the vendor for modifications, ease of data entry in real time, ease of data retrieval in real time

- **Easy support for capturing unique client records**
  - a. Automatically generated Unique Client Identifiers
  - b. Key demographic data
  - c. Key baseline data
  - d. Key outcome data
  - e. Service utilization (including premature exit from services and reasons why)

- **Support for front-line staff**
  - a. All required data collection forms are incorporated (no more pencil and paper forms)
  - b. There is adequate space for recording notes and other qualitative information
  - c. Easy to use tools support staff in managing relationships with clients – e.g., planning/scheduling of meetings, recording/tracking client contacts
  - d. Client service utilization is tracked
  - e. Client progress is tracked
  - f. Staff efforts/activities are tracked

- **Support for supervisors**
  - a. Tracks staff efforts – individually and aggregated by caseload, program, division, job category, etc.
  - b. Tracks client progress toward stated objectives – individually and aggregated by caseload, program, division, job category, etc.
  - c. Tracks client service utilization – individually and aggregated by caseload, program, division, staff job category, etc.

- **Support for higher-level management needs**
  - a. Generates aggregated Enrollment Reports – across all programs, divisions, etc.
  - b. Generates aggregated Client Demographic and Baseline Profiles – across all programs, divisions, etc.
  - c. Generates aggregated Client Service/Program Utilization Reports – across all programs, divisions, etc.
d. Generates aggregated Client Outcome Reports – across all programs, divisions, etc

e. Provides tools for routine as well as customized Data Analysis – across all programs, divisions, etc

◆ Report generation – degree of flexibility for end users at all levels of the organization to design/generate reports

◆ Web-based – supports remote, flexible access by all end users
  
a. Hosted by the organization on its own server(s)
  
b. Hosted by an Administrative Service Provider (ASP)

◆ Architecture and scalability
  
a. Degree to which system can support large organizations – how much simultaneous use by large numbers of end users entering and retrieving data affects system performance
  
b. Ease with which system can be implemented and used across multiple sites (nearby and distant)

◆ Implementation – ease with which system can be installed and configured (including the use of each organization's unique terminology)

◆ Training – Availability, quality, and cost of staff training by vendor (both at time of implementation and at periodic follow-up intervals)

◆ Security
  
a. Internal (protected levels of access within the organization, tracking history of individual records and their modification with audit capability)
  
b. External (protection against external access to the organization's data, redundancy, remote daily automated back-up, ease of data retrieval in cases of disastrous system events)

◆ Migration tools – for uploading and cleansing historical or legacy data; such tools can make it possible for organizations to undertake this themselves (as opposed to the vendor doing so at added cost)

◆ Integration capabilities – ease of transferring data to/from other systems (e.g., funder reporting systems) on an ongoing basis (often requires custom programming by vendor at added cost)

◆ Technical support – cost and availability of on-site and remote technical assistance over time
David E.K. Hunter, Ph.D.

Since September 2006 David Hunter has been working independently, both in the United States and internationally, consulting to the social (not-for-profit) and public sectors with a focus on organizational capacity building, developing strategies and theories of change, performance management, and the creation, delivery, and assessment of social value. His practice builds on over three decades of experience, in both sectors, using performance management systems to improve the quality and effectiveness of social services. As the Superintendent (CEO) of a State Psychiatric Hospital in Connecticut, his efforts led to improved hospital safety and lower patient length of stay while attaining measurable, good treatment outcomes, and led to a rating of Accreditation with Commendation by the Joint Commission on the Accreditation of Health Care Organizations (JCAHO).

Subsequently, as Director of Evaluation and Knowledge Development at the Edna McConnell Clark Foundation in New York City, David worked with nonprofit organizations to develop clear value propositions through capacity-building “theory of change workshops,” and to design, implement and use performance management systems to deliver, monitor, learn from, and evaluate high quality, effective, and efficient human services. Dr. Hunter is the author of numerous articles and papers concerning strategic performance management and how to create, invest in, and sustain social value in complex situations with diverse stakeholders. He can be contacted at: david@dekhconsulting.com.