
Defining the Bottom Line:
An Evaluation of the Black Ministerial Alliance's
Performance Management Capacity Institute Pilot

January 31, 2013

By Carl Sussman, Sussman Associates and
Kim Comart, Comart & Associates

Contents

Executive Summary of Key Themes and Findings	2
Evaluation Methodology	3
Theory-Based Evaluation	3
Developmental Evaluation	4
Capacity Building Evaluations	4
Evaluation Questions	4
Data Sources	5
Organizational Capacity Building Context	6
Findings and Recommendations	8
Overall Assessment of Impact	8
Participant Selection Process	9
Project Duration	9
Cohorts and Communities of Practice	10
Organizational Readiness	11
Benchmark Assessment Tool	12
Technical Assistance Work Plan	17
The Capacity Building Process	18
1. Theory of Change Workshop	18
2. Group Training Sessions	20
3. Individualized Technical Consulting	22
4. Coaching	22
References	24

Executive Summary of Key Themes and Findings

The Capacity Institute is noteworthy because the performance management framework represents a comprehensive approach to outcome measurement, improving program effectiveness and strengthening organizational infrastructure. We found compelling evidence that the Capacity Institute grantees engaged in an intentional and intensive process of reshaping their programs to be more results-focused and their management practices to reflect the principles of performance management.

To our knowledge, previous outcomes-oriented capacity building programs have not been as comprehensive nor as successful in addressing the many barriers organizations face in implementing these systems and using them to produce better performance and outcomes. To be successful, these efforts must be tied to transformational changes in an organization's culture as well as a more comprehensive approach to the design and implementation of the data collection-analysis-accountability-program improvement cycle.

Based on our review of the BMA's extensive documentation of its activities, interviews and in-depth monitoring of the three case study organizations, we feel the Capacity Institute's model has addressed weaknesses that were apparent in previous efforts to use an outcomes orientation to organizational and programmatic capacity building. Thus, the Capacity Institute represents a very promising model.

Because of the pioneering nature of this program design, we adopted a developmental approach to the evaluation – not only assessing the final results, but placing special emphasis on identifying lessons from the initial implementation that can help the Black Ministerial Alliance (BMA) strengthen the program for future cohorts. Among these lessons are:

- **Readiness** – All of the participating organizations appeared to benefit from the program. The BMA did an especially good job in selecting organizations it could motivate to stick with the Capacity Institute. At the same time, we believe a tiered approach should be considered for future cohorts, one that provides more intensive assistance to and expects more from the larger organizations with greater capacity. Smaller organizations and faith-based programs that rely primarily on volunteer staffing lack the capacity to consume and take advantage of the full menu of technical assistance offerings provided by the BMA. These organizations would benefit from a more targeted and customized approach that is both realistic and measures their progress by a different yardstick.
- **Length** – One significant question in any program evaluation is dosage – just how much activity is needed to achieve the desired outcomes. Given the type of profound organizational change that is involved in adopting a performance management system and culture, the program should at a minimum have a longer tail, where the organizations receive on-going monitoring and implementation support.
- **Cohorts** – The shared focus on youth-serving organizations proved to be one of the Capacity Institute's strengths. However, because of the diversity of the participating organizations there probably was not enough similarity to achieve the type of synergies that peer-to-peer group activities can generate. In addition, the BMA might consider creating more structured

opportunities for peer-to-peer learning. One option would be to form one or more communities of practice of the participating executive directors -- perhaps segmented by size -- and arrange for them to meet regularly. It might be useful to form a similar group of program directors. In any event, this is one way to ensure ongoing peer support after the Capacity Institute assistance formally ends.

Evaluation Methodology

This evaluation uses a theory-based developmental approach to program evaluation. This methodology is a form of action research. It is designed to document what the grantee and its collaborators did; figure out what worked (or didn't work); provide informed thinking about the factors that contributed to or undermined success, and, supply independent perspectives and information to strengthen the program theory and execution. The purpose of the developmental aspect of this Capacity Institute evaluation is not primarily to cast judgment about the program operator's performance; rather it is to promote learning and improvement since both the funders and grantees share the same goal: To figure out effective ways to strengthen performance management in community-based organizations serving high-risk youth in Boston.

As with any evaluation, the design reflected difficult resource allocation trade-offs and implementation constraints. First, implementation time pressures meant that the data collection began after some important project-related activities had already occurred, such as the recruitment and selection of organizations to participate in the project and the design of the benchmark tool that would be used to conduct the baseline and final assessments of the participating organizations. Second, the project budget was extremely tight and therefore the resources devoted to the evaluation were severely limited. Nonetheless, while we would have preferred to have started earlier in the process and had the time to monitor more organizations and observe more of the activities, the evaluators believe enough information has been captured to assess the degree to which the program achieved its goals and to formulate recommendations for how the program design could be strengthened.

Theory-Based Evaluation

According to Carol Hirschon Weiss, theory-based evaluation gauges a program "against the explicit claims and tacit assumptions that provide its rationale." Theory, in this case, means "the logical series of steps that lays out the path from inputs to participant responses to further intervention to further participant responses and so on, until the goal is achieved (or breaks down along the way)." The theory-based approach evolved to cope with the challenge of evaluating interventions occurring in inherently complex circumstances and within open systems; conditions in which it is impossible to construct a controlled experiment. As Weiss explains,

In these cases, evaluators want some way to try to attribute causality. They want to be able to say, "The program caused these outcomes." Without randomized assignment, causal statements are suspect. But if evaluators can show that the program moved along its expected sequence of steps, and that participants responded in expected ways at each step

of the process, then they can claim a reasonable approximation of causal explanation.
(Weiss, p. 3).

Developmental Evaluation

Developmental evaluation is a relatively new branch of evaluation research. It has emerged to meet the distinctive needs and characteristics of untested programs. Formative and summative evaluations are most useful in documenting and assessing stable operating models. “When a formative or summative evaluation approach is applied to an innovation that is still unfolding, it can squelch the adaptation and creativity that is integral to success.” (Preskill & Beer, p.5). By contrast, the purpose of developmental evaluation is to support continuous learning and stimulate program adaptation.

Because the developmental evaluator serves as “a strategic learning partner and facilitator,” the relationship with the BMA is somewhat different from the more detached relationship that characterizes other evaluation models. (Preskill & Beer, p.8). For example, the evaluators jointly planned with the BMA, and then facilitated a learning session with program participants near the conclusion of the project. So, in addition to contributing evaluative thinking, data and analysis to the project team’s implementation efforts, the developmental evaluator may also serve as a facilitator for, or even consultant to, the program operator.

Capacity Building Evaluations

The Capacity Institute’s ultimate goal is to improve youth outcomes. This evaluation only covers the initial and intermediate outcomes listed in the Institute’s logic model, most of which relate to mission clarification, program design, outcome measurement, organizational infrastructure, and internal alignment. Outcomes for youth are understandably long-term; they are expected to occur down the road as a result of the participating agencies’ implementation of their performance measurement systems.

In any event, an evaluation research design capable of documenting such long-term outcomes would be extremely costly, would need to follow these organizations (and the youth) for many years and would pose other major challenges. (Simister, N. 2010). This problem of linking capacity building with the long term outcomes they are designed to achieve bedevils all capacity building evaluations. Paul C. Light, in a meta-evaluation of capacity building initiatives, has written “...capacity-building outputs must eventually find their way to actual program impacts....Unfortunately...the linkages between activities, outputs and outcomes are often anything but clear.” (2004, p 89). In the end, relying heavily on subjective data, Light concludes that organizational capacity makes a critical difference in mission effectiveness. However, the value of capacity building efforts to specific organizations is situational, especially with respect to an organization’s stage of development.

Evaluation Questions

This evaluation attempts to answer the following eight questions:

Defining the Bottom Line

1. How successful were the organizations in accomplishing the tasks (developing a logic model, defining outcome indicators, developing a database, etc.) required to develop a performance management system?
2. To what extent have the 17 organizational practices been built in the participating agencies?
3. Did the participating organizations take meaningful productive steps to improve their outcomes using performance management?
4. How likely is it that improvements that have been made in the organization through this initiative will be sustained over the next year? The next 2-5 years?
5. Is the organizational benchmark assessment tool a valid measure of high-quality organizational practices and organizational strength?
6. What elements of the Capacity Institute were most effective in advancing the goals of the project? What elements were least effective?
7. How effective was the project's implementation? What were the ramifications for the project's effectiveness?
8. What lessons can be learned to inform replication of the project?

Data Sources

The evaluator collected data from a variety of sources. These included:

1. **Participant observation** – The evaluator attended a small sample of program activities, including a theory of change session, consultant-led technical assistant consultations and workshops.
2. **Monitoring information flows** – The BMA provided the evaluators with voluminous documents relating to the Capacity Institute and each of the organizations.
3. **Periodic consultations** – Members of the project team made themselves available to meet in person or by phone for periodic updates on programmatic activities.
4. **Case Studies** – The evaluators selected three participating organizations to be followed throughout the process. Using program documentation, participant observation and interviews, the evaluators were able to gain insights into the changes that occurred in organizations during the program and gauge the impact of the program. It was agreed that these would not be presented as case study narratives for this report because of resource limitations.
5. **Key Informant Interviews** – The evaluators conducted a series of key informant interviews with key participants in the program to collect qualitative data and different perspectives on the activities.
6. **Focus Groups with Funders and Grantees** – The evaluators met with staff from the foundation funders to solicit input regarding their expectations for program outcomes and the type of evaluative information they were interesting in seeing. They conducted two similar data gathering sessions with the agencies participating in the Capacity Institute.

Organizational Capacity Building Context

It is important to frame the Capacity Institute in the context of other purposeful efforts to strengthen organizations in the nonprofit sector. While foundations often describe their philanthropic grants as “investments,” in most cases they actually “buy” direct services rather than “build” the grantee’s productive capacity. The latter, by generating future productivity, conforms to the conventional distinction between expenditures and investment. Capacity building is a form of investment. Typically grants and government contracts don't pay to strengthen the support activities in the organization’s value chain: Human resource management, financial management systems, office automation or improved physical infrastructure (Overholser). This is certainly true for most nonprofit human service organizations, such as community-based youth development programs. These are “widely (and correctly) regarded as indispensable, yet,” as Clara Miller has noted, their “underlying economics and management systems seem to be almost no one's primary concern.” (Miller, p.45). This has begun to change.

The 1998 publication of *High Performance Nonprofit Organizations: Managing Upstream for Greater Impact* (Letts, C.W., Ryan, W.P. & Grossman, A.) helped crystallize a critique about the sector’s failure to have a greater impact on a wide range of critical social challenges. The authors exposed the deeply engrained bias within the nonprofit sector – among both philanthropic and governmental funders as well as the nonprofit organizations they fund – to spend money and effort in programs (buying services) while ignoring the need to invest in the underlying organizations. The book makes the case that well-conceived programs can be expected to fall short of expectations without high performance organizations driving their implementation. The authors’ findings have become conventional wisdom and helped spur a growing appreciation for the importance of organizational “capacity building” – direct efforts to enhance organizational productivity, effectiveness and sustainability. Investments in organizational capacity building strengthen a nonprofit’s operations by building the internal systems and structures appropriate to the size, complexity, and stage of growth of the organization. As Venture Philanthropy Partners summarized the rationale for building organizational capacity in *Effective Capacity Building in Nonprofit Organizations*:

All too many nonprofits...focus on creating new programs and keeping administrative costs low instead of building the organizational capacity necessary to achieve their aspirations effectively and efficiently. This is not surprising, given that donors and funders have traditionally been more interested in supporting an exciting new idea than in building an organization that can effectively carry out that idea. This must change; both nonprofit managers and those that fund them must recognize that excellence in programmatic innovation and implementation are insufficient for nonprofits to achieve lasting results. Great programs need great organizations behind them. (McKinsey & Company, p. 39)

Another important development in the nonprofit sector during the mid-1990s was United Way of America’s initiative to alter the framework for thinking about program effectiveness. Until then nonprofits primarily tracked financial performance, service delivery, participant demographics and adherence to accreditation and other quality standards.

Defining the Bottom Line

Increasingly, however, there has been recognition that, while such measures show how much effort has been generated for how many individuals, they reveal nothing about whether this effort has made any difference – whether anyone is better off as a result of the service. Outcome measurement responds to this gap. (Plantz, Greenway & Hendricks (2006) p. 3).

United Way of America was a pioneer in the movement towards an outcomes-based approach to program accountability. It reflected a dramatic paradigm shift from reporting activities to measuring results. The shift to outcomes measurement not only added a new level of accountability; it also served as a lever for driving program improvement as organizations used outcomes data to inform the design and implementation of their programs in order to achieve better outcomes. The literature and practices gradually became more sophisticated, absorbing performance management practices used in the private sector, as reflected in a raft of books including Mark Friedman's *Trying Hard is Not Good Enough* and Mario Morino's *Leap of Reason: Managing to Outcomes in an Era of Scarcity*. Another book on this subject, by Capacity Institute consultant David Hunter, will be published in February, 2013.

As we reflect on the nonprofit capacity building movement over the past two decades, we see that it has evolved in three phases. The first involved investments in formal organizational structures and procedures – board and leadership development, strategic planning, financial management systems, technology infrastructure, and fund development capacity. The second centered on reorienting organizations to outcomes. While the first phase focused on building the capacity of the corporate entity, the second phase was programmatically focused. Capacity building initiatives included organizing training workshops and assembling cohorts of organizations to construct logic models and measurement systems. The capacity building focus of this second phase has been on outcomes reporting, primarily for accountability purposes. The third phase involves the internalization of outcomes measurement and its unification with major elements of the first phase of the organizational capacity building movement. This phase has seen nonprofit organizations embrace performance management as a tool for advancing their mission. Organizations that are building their capacity for performance management appear to be at least as internally as externally motivated and more interested in clarifying and uniting programmatic outcomes with a coherent organizational mission and strategy. The most important aspect of the performance management capacity building phase is the deep cultural change it entails. While the outcomes measurement phase posed what Ronald Heifetz described as a technical challenge, the performance management phase involves the much more difficult challenges associated with adaptive change. (Heifetz & Linsky. 2002). The Capacity Institute is in the vanguard of this third phase.

Findings and Recommendations

Overall Assessment of Impact

As a developmental evaluation, this document focuses on areas where the Capacity Institute could be made more effective. However, it is important to emphasize the amount that has been accomplished and the organizational transformations it is responsible for causing. One of the consultants interviewed for this evaluation stepped back at the end of the interview, reflecting on the analytic details that consumed most of the interview and said:

It has been fascinating to see where the organizations were and are now. The overall project is pretty amazing. For some of the organizations, they had no understanding of how to track what they were doing. [One organization] tracked all this amazing information on paper. They had no understanding of the power of that until they went through this process. The idea of tracking the data and evaluating whether the program is effective is an extremely powerful thing. Over the course of two years some of these organizations went from “we think we are doing well” to “here is how I can make it better and here is how I could pivot to be more effective in my community.” I heard this from many organizations. Many of the organizations had been creating programs based on funding, not mission.

Without exception, this was the overriding sentiment reported by everyone interviewed for this evaluation. This was true even for the few organizations that were least engaged and least ready to benefit from the program. The Capacity Institute was audaciously ambitious and, despite many constraints, it achieved transformational results for some of the organizations and dramatic changes in programmatic thinking and action for just about all of them. As evaluators who have provided management consulting services to scores of nonprofit organizations, we have concluded that the Capacity Institute was a remarkably cutting-edge and successful organizational capacity building project.

Success in this case means that there is compelling qualitative evidence that the organizations have become more focused on outcomes and committed to using data to improve them. They have begun to build the systems and apply practices that will enable them to use data to generate better outcomes for youth. And there is a greater sense of organizational and individual accountability for achieving outcomes although most of the organizations have yet to implement or fully implement their outcome measurement systems and reorient their human resource functions to support that culture change.

The big outstanding question is how much of this change will stick. To what extent will these organizations continue to build on the systems they are constructing? In other words, has this project really initiated a sustainable change in organizational culture that will lead to improved performance? It is too early to draw any conclusions. Given the forces working against such change, it is inevitable that some of these organizations will backslide. However, based on our interviews and observations, we believe it is equally possible that many of these organizations will be permanently changed for the better. On this score we can assert that the Capacity Institute is indeed a promising model.

Participant Selection Process

The Capacity Institute was a two year project. Recruitment of participants began in December of 2010 with the dissemination of applications which were due in mid-January of 2011. While the BMA hoped to begin the recruitment in October, when the Corporation for National and Community Service grant began, it needed to wait for federal government approval of the application form, which did not occur until early December. There were 10 applicants in that first round; of those, five were selected. They were notified in early February. Shortly afterwards, one was declared ineligible, leaving four organizations in the pool. A second-round deadline in mid-March yielded four applications. The BMA approved all four. Finally, eight organizations applied for a third round in early April. Six were accepted, bringing the total Capacity Institute cohort to the target number of 15.

Capacity Institute program manager Ellen Bass worked hard to recruit organizations to participate in the project. She found the task surprisingly challenging. Organizations were resistant for different reasons. It was clear that the program would require a very significant commitment of time and organizational resources for up to two years, including many hours of the executive director's and key program staff's time, as well as some level of participation by the Board of Directors. While each organization received a token cash grant of \$2,000 and consulting and training valued at approximately \$25,000, the opportunity cost in staff time more than offset the financial gain. Moreover, this was a new project with no proven track record of success that prospective participants could factor into their decision about participation. Finally, Bass made it clear that hard questions would be asked and traditional ways of doing business would be challenged. Engagement in the Capacity Institute was not for the faint of heart.

Recruiting organizations to participate in future rounds of the Capacity Institute may become easier. The BMA will be able to point to a successful first phase of work, with lots of evidence of success and personal testimonies. In addition, as other youth-serving organizations become aware that their peers have strengthened their performance management thru the Capacity Institute, they will feel some pressure to do this work themselves in order to compete for funding and recognition. Finally, Bass will be able to articulate the benefits of participation and make a convincing argument for entry into the program.

Project Duration

Slated to begin on October 1, 2010, the Capacity Institute ramped up in earnest over a six-month period from January to June 2011, with three cohorts of organizations joining during that period. Over time, some of those organizations dropped out and others joined, with the last entry in January 2012. As such, most of the organizations ended up being in the program less than 18 months; for a few it was closer to 12 months or less. This is in contrast to the original project design, which envisioned a full two-year capacity building program.

The Capacity Institute was appropriately intensive. It had to be. However, there are two issues that have fascinated us about the dosage. One is duration. The consensus opinion is that it needed to be longer. But how long? The other is how it should end. Does it end at a cliff or should it taper off more gradually

and, if so, how gradually? Given the challenges of initiating and sustaining change in an organization's culture, both of these issues are important.

The most common answer we heard about length was that it needed to be six months longer. Ironically, for some organizations that means a program that would be 18 months long, whereas for those who began the process earlier, six more months would mean the two year period the project was designed to be. We agree that the Capacity Institute should have been at least 24 months *from the point at which organizations were accepted into it*. However, we suspect performance management will be an ongoing capacity building challenge for all organizations. So the more important issue is how to support their capacity building on an ongoing basis.

Despite the fact that the project ended at in September, the BMA continues to assist many of the organizations and has committed to quarterly check-ins with each of them. We think it would be advantageous to be able to continue to provide coaching at a reduced level for as much as two years while the organizations pilot test and make appropriate improvements to their data collection systems, and while the participants fully implement their redesigned human resource systems and procedures.

Cohorts and Communities of Practice

Another less staff-intensive way to create ongoing support is by more intentionally forming and cultivating communities of practice among groups. In fact, formalizing cohorts as communities of practice at the outset might provide additional impetus to sustain organizational commitment to performance management, both during and following the Capacity Institute, as well as providing a great vehicle for problem solving and introducing new staff to the performance management culture.

Communities of practice, according to Wenger, McDermott and Snyder, "are groups of people who share a concern, a set of problems, or a passion about a topic, and who deepen their knowledge and expertise in the area by interacting on an ongoing basis." These authors describe communities of practice this way:

As they spend time together they typically share information, insight, and advice. They help each other solve problems. They discuss their situations, their aspirations, and their needs. They ponder common issues, explore ideas, and act as sounding boards. They may create tools, standards, generic designs, manuals, and other documents – or they may simply develop a tacit understanding that they share. However, as [sic] they accumulate knowledge, they become informally bound by the value that they find in learning together...They may even develop a common sense of identity. (p. 4-5).

Communities of practice occupy the intersection between a "domain of knowledge, which defines a set of issues," in this case issues relating to youth development; "a community of people" who care about a topic, and "the shared practice [performance management] that they are developing." (p. 27). Indeed, at the final meeting convened to assess the Capacity Institute experience, one executive director expressed the desire to spend "more time with other agencies, similar agencies, to learn from each other. What did they learn that might influence us?" Another director observed that she "Liked that we

got to hear what [other] people were going through. [We] wanted to hear more about what people are going through. I learn from hearing specific things. I'd be willing to share more. Maybe people get a partner, paired for the duration." We think both of these comments reflect both the sense of community that exists among these organizations and the desire to have a learning community; something that could be a resource during the Capacity Institute but especially after it.

Organizational Readiness

As with any technical assistance program, the ability to reap the benefits requires sufficient levels of organizational readiness and commitment. That means motivation to commit to the work – to make change among leadership, program staff, board and key volunteers --and the organizational capacity in terms of programmatic sophistication, and a sufficiently large staff to have some degree of specialization and among whom to distribute the work.

Interviews with Bass and the Capacity Institute consultants, and the participant organizations' staff and funders, reinforced the importance of organizational readiness. For most of the participants, organizational motivation was strong; they were prepared to dive in with both feet and had the capacity to fully engage and stick with the program. On the other hand, a few of the organizations were not ready. Among the valuable lessons gleaned from this pilot implementation of the Capacity Institute is that these few organizations should not have been included in the program. One faith-based organization withdrew when its leaders realized that their religious values were in conflict with the federal government funding restrictions. Some smaller organizations struggled because staff lacked the necessary bandwidth to absorb the Institute's additional time demands. A few large and diversified organizations encountered internal challenges because the senior leadership was not as committed to the project as program managers were although this did not prevent them those organizations from making important gains.

The three key elements of organizational readiness were:

- **Commitment of the organization's top staff and board leadership** – When the executive director was fully engaged and the board was fully informed and supportive, the likelihood of progress increased dramatically. If they were not fully engaged, the organization was far less likely to reap the full benefits of the Capacity Institute, despite the fact that there may have been other enthusiastic members of the staff. Bass included leadership among a handful of "deal breakers" – barriers to the organizational culture change that is requisite to implement and sustain a performance management regime. She had to challenge six organizations to resolve resistance to or ambivalence about the changes needed to instill a performance management culture.
- **Organizational size and capacity** – Regardless of the level of enthusiasm for participating in the Capacity Institute, all other things being equal, those organizations with larger budgets and more staff were more likely to make progress and are more likely to sustain the momentum after the Institute. As noted elsewhere in this report, the Capacity Institute demanded a great deal from the cohort members, particularly staff time. Board members were expected to

Defining the Bottom Line

participate in some of the sessions. Several of the small organizations with only two or three paid staff found themselves stretched too thin to be able to take full advantage of the resources and keep pace with the Institute timeline. Also, organizations with more personnel are likely to have more specialization and role differentiation, improving the likelihood that the organization could manage the Institute's time and content demands.

- **Organizational sophistication** – Cohort members entered the program at different levels on the performance management continuum. Some had more familiarity than others with the outcomes orientation and had embraced the performance management mindset even if they didn't have rigorous systems in place. Organizations with effective leadership, engaged boards, professional development activities, and other signs that they were receptive to change are good candidates for this Initiative. For those participants, the Capacity Institute represented an opportunity to leap forward and become very high-functioning non-profits. For organizations starting from a very low level of organizational sophistication, the Capacity Institute presented an entirely new paradigm and a huge climb.

The cohort included organizations that were ready and for whom the Institute was transformative. They stuck with the project and made major gains. A few members of the cohort proved to be less ready than they may first have appeared and at best they made limited progress.

If and when the BMA replicates this pilot, among the changes it might consider are changes to the selection process to better assess each organization's readiness. The BMA might augment the application with in-person interviews with the executive director, other senior managers, and a mix of board members to assess their openness to change, their commitment to performance management, and their willingness to raise funds to support the performance management infrastructure – technology, training and staffing – to maintain the investment.

Benchmark Assessment Tool

Practicing what it preaches, the Capacity Institute developed a data collection tool for assessing each participating organization's capacity for performance management as it existed at the outset of the project and again at the end. Each agency had to use the Capacity Institute Benchmark Assessment Tool to self-assess its performance management capabilities against 17 benchmarks in four domains:

- Utilization of a positive youth development framework
- Measures program outcomes and uses them as a vehicle for learning and improving programs
- Fully documents its program designs and communicates transparently with funding partners about performance
- Development of partnerships and coordination of services

After completing the assessment, the organization's leaders met with Ellen Bass, the Capacity Institute's director, to review and discuss the basis for the scores. As a result of these reviews, each of the organizations adjusted some of the scores. The procedure was repeated at the conclusion of the grant and the scores were compared to gauge improvement.

Defining the Bottom Line

The basic framework copies other organizational capacity rating tools such as the *McKinsey Capacity Assessment Grid* (2001). A successor assessment tool, the *Marguerite Casey Foundation Organizational Capacity Assessment Tool* (2002) appears to have served as the Excel template for the Capacity Institute Benchmark Assessment Tool. Associated with each of the Capacity Institute's benchmarks are four descriptors. Level 1 descriptors depict an absence of the practice whereas level 4 captures an organization's mastery of the practice. For example, the first level for Benchmark 3, ("Agency completes a logic model which clearly and concisely demonstrates its unique theory of change") describes the circumstance where the organization has "no program logic model, or no youth outcomes articulated." By contrast, at level 4 there is a "logical, tight flow of influence from activities through the entire sequence of outcomes, leading to meaningful long-term outcomes; logic model elements are stated clearly, comprehensively, and concisely."

One question we were asked to address was whether the tool is a valid and reliable instrument. Reliability is a measure of consistency. Anyone with sufficient knowledge of an organization who uses the instrument to rate the same capacity will arrive at a very similar value. A tool can be reliable but not valid. Validity refers to whether the tool actually measures what it purports to measure. To use an absurd example, you cannot measure someone's height with a thermometer. Ninety-eight degrees would not be a valid value for a six-foot tall person.

We did not design an experiment to test the reliability and validity of the tool, in part because it was designed and used for self-assessment. We found evidence that it is not entirely reliable *as used*. However, if put to the test, the tool may prove to be reliable when used by trained third party raters, as most observational rating tools are. One of the organizations interviewed noted that when they compared their assessment of their capabilities at the end with their initial assessment, they felt in hindsight that they had unintentionally inflated their initial scores. In effect, the organization conducting the assessment the second time was not the same rater: They had a better understanding of performance management and were more sophisticated in their judgments at the conclusion of the project.

While we neither devised nor conducted an experiment to establish the tool's empirical validity, we are unaware that any of the nonprofit organizational capacity assessment tools following the same general structure and developed over the past ten to twelve years have been subjected to empirical scrutiny. However, it is our opinion that a strong case can be made for the tool's face validity, based on our experience in organizational assessment and performance management. The scope of the tool – the domains and the specific benchmarks – is reasonably comprehensive and the gradations (levels) within each benchmark describe logical developmental steps that lead to the level four benchmarks.

Most importantly, the benchmarking tool proved valuable to participants in the program because it provided a map of goals they were trying to achieve; a sense of where they stood; feedback about progress, and goals to strive towards.

We think the Capacity Institute's tool does a good job of creating descriptors that make it possible for a person with appropriate experience or training to arrive at a reasonably objective score. We have two

Defining the Bottom Line

suggestions for improving the reliability of the tool. First, separate the descriptions that have multiple properties and provide the instruction that all must be true to earn a score at the specific level. For example, for level two on benchmark twelve, you might reformat the description like this:

Program plans accurately reflect:

- staff activities**
- description of goals, objectives and clients, and**
- outcomes are moderately complete and clear**

Second, the BMA should consider supplementing these descriptions with examples. For instance, the level three and four descriptions for benchmark 10 might give an example of what would constitute “agency uses outcomes data regularly to improve service delivery:”

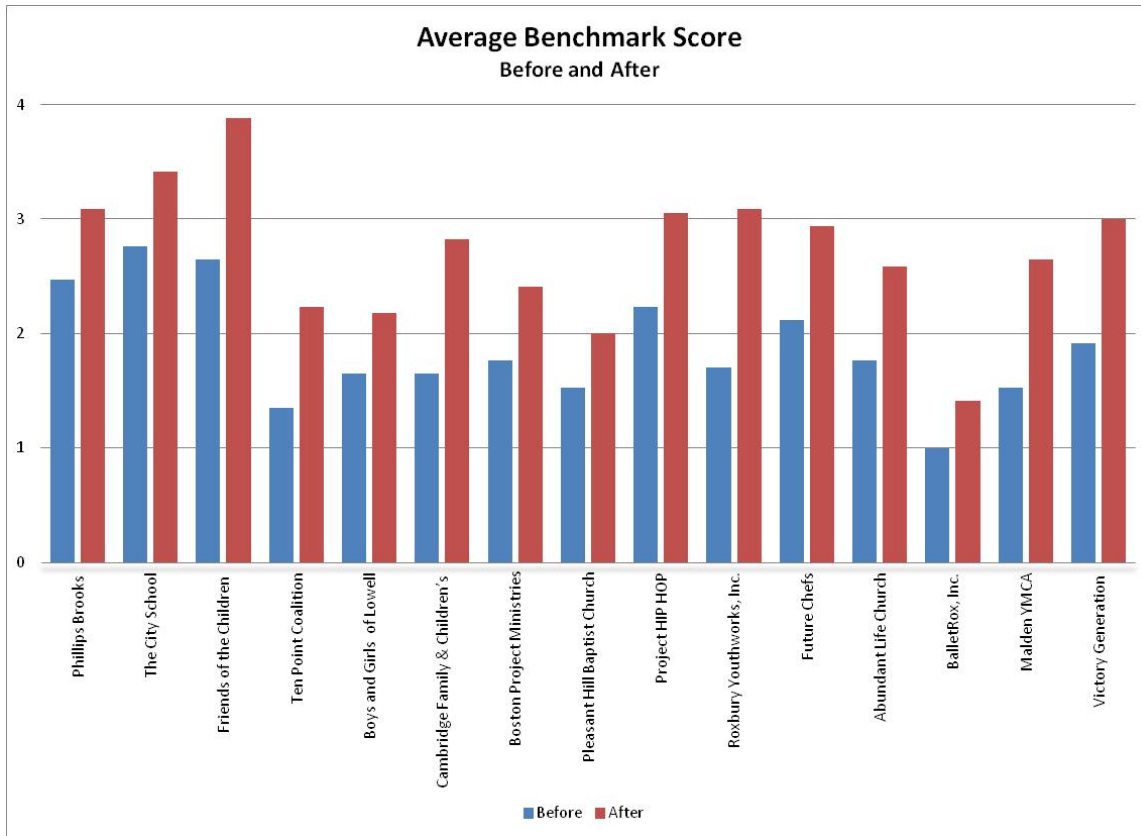
Example: Outcomes reports are prepared at least quarterly AND team meetings with management are convened to discuss these reports and, where necessary, consideration is given to change a program’s design, procedures, resourcing or staffing to improve outcomes.

The results of the benchmarking exercise, as seen on the graph on the top of the following page, proved that the organizations started at very different points in terms of the performance management capacity. And all showed measurable improvement based on the benchmarking assessment process. While it is impossible to confirm the absolute values or even the relative values between organizations, the interviews confirmed that all of the organizations had made obvious gains.

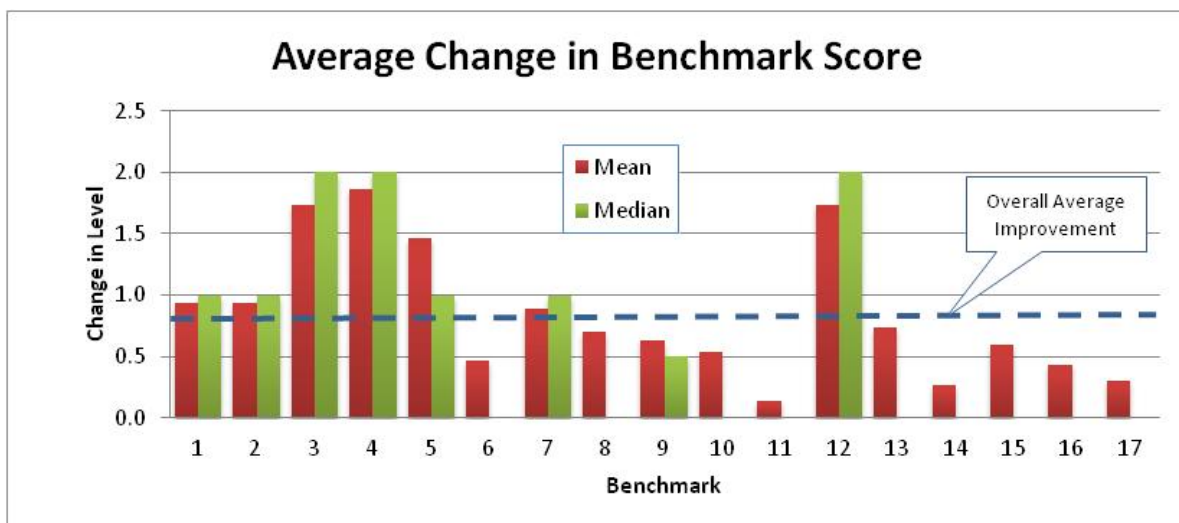
In general the larger agencies appeared to have improved more on the benchmarks. These organizations had more resources to devote to the project in terms of staff who could be assigned, and the staff they had tended to be more specialized, again as a result of organizational scale. One person speculated that the smaller agency had an easier time with changing the culture because there is less built-in resistance. On the other hand, small organizations may find it more difficult to build the organizational systems needed to support and sustain new practices.

The change in benchmark scores, when taken as a whole by looking at the average change in score, reveals some interesting patterns. The following graph at the bottom of page 15 shows two measures of central tendency. The mean is the conventional understanding of average – the sum of all the changes divided by the number of organizations. The mean is shown in the following graph in red. Median is a more telling measure of central tendency when you have a skewed distribution. The median is shown in green. Looking at the mean change, the organizations on the whole experienced improvement (>.5 but <1.5) for approximately half of the benchmarks. They showed dramatic improvement (>1.5) on average for three benchmarks:

Defining the Bottom Line



3. Agency completes a logic model which clearly and concisely demonstrates its unique theory of change.
4. Agency identifies specific indicators to measure youth outcomes, an initial and an intermediate outcome.
12. Agency develops written plans for each service delivery area: goals, objectives, target clients, projected outcomes.



Defining the Bottom Line

The first two of the benchmarks demonstrating dramatic improvement (3 and 4) were also the ones in which the Capacity Institute invested the most effort as measured in consulting assistance. (See the following table).

Capacity Institute Total Hours to Benchmark Practices

Agency	Benchmark Number														Total
	1	2	3	4	5	6	7	8	9	10	11	12	13	14-17	
Phillips Brook House Assn.	2	31	21	8	33	17	0	0	19	2	0	0	0	16	149
Malden YMCA	10	10	10	3	6	16	4	0	3	0	0	0	0	16	78
Boston Project Ministries	2	25	20	7	14	16	0	0	19	0	0	0	0	21	124
Boston TenPoint Coalition	2	26	24	18	10	0	0	0	8	2	0	6	0	20	116
The City School	2	22	16	13	16	34	0	2	18	4	0	0	0	16	143
Pleasant Hill Baptist Church	10	13	36	9	2	16	1	0	6	11	0	0	0	16	120
Project HIP HOP	2	20	18	12	5	16	1	0	18	0	0	0	0	17	109
12th BC/GRASP Inc./VG	2	42	11	8	1	19	4	7	17	0	0	0	1	0	112
Roxbury Youthworks, Inc.	2	22	18	14	20	34	0	0	18	5	0	2	1	0	136
Friends of the Children Boston	2	20	9	11	5	3	1	0	19	0	0	0	1	17	88
Future Chefs'	2	20	19	4	0	19	0	0	0	0	0	1	0	16	81
Cambridge Family Children Serv	50	29	19	17	1	3	0	0	14	9	0	0	1	16	159
BalletRox, Inc.*	2	30	12	2	1	0	0	0	8	45	0	1	0	0	101*
Boys & Girls Club of Lowell	2	20	27	0	0	1	0	0	8	0	0	0	0	0	58
Abundant Life Church	2	20	1	7	4	16	0	0	2	0	0	2	0	0	54
Average all agencies	6.3	23.3	17.4	8.9	7.9	14.0	0.7	0.6	11.8	5.2	0.0	0.8	0.3	11.4	108.5

NOTE

* BalletRox's hours on data use were really preliminary for benchmark 10 goal of leadership development

Much less significant improvement (< .5) occurred on five of the benchmarks, although two of them, benchmark 11 (agency demonstrates improvement in youth outcomes over time) and 17 (agency develops partnership with schools and/or employers to ensure youth outcomes) are long-term outcomes projected to occur sometime after the two-year project period. That leaves three benchmarks where participating agencies on average showed less significant improvement:

6. Agency has technology systems in place to track indicator data for each outcome.
14. Agency develops and uses common youth assessment and referral protocols with service partners.
16. Agency refers enrolled youth to appropriate partners for relevant services.

All three of these outcomes appear as initial outcomes in the Capacity Institute's own logic model. However, sequentially, the Capacity Institute's partnership and technology systems components were the last introduced to the cohort of agencies. As pointed out earlier, while the project was designed as a two year intervention, because of unavoidable start-up delays and unpredicted challenges recruiting organizations that met the eligibility criteria, the project activities occurred over 18 or fewer months. Thus, it is not surprising that these were the outcomes where the project saw the least improvement. Also, again using hours of consulting time as a measure of investment in benchmark improvement, benchmarks 10-11, 13-17 received the least investment. (Refer to the preceding table).

The Average Change in Benchmark Score graph also displays the median improvement for each benchmark. Divergence between the mean and the median for any one of the benchmarks means that the distribution was skewed. The most skewed distributions are for benchmarks:

10. Agency uses outcomes data to improve service delivery, reward staff performance, and inform management decisions.
13. Agency regularly informs stakeholders and funders of its youth outcomes.

Defining the Bottom Line

15. Agency refers youth it does not serve to appropriate partners for relevant services.

These three outcomes were identified as intermediate outcomes in the Capacity Institute’s logic model. The lack of progress for the large proportion of agencies in these three cases can be attributed to the sequencing of activities and the shortened performance period. However, it is noteworthy that the skewed distribution also means that despite these constraints a small number of agencies made significantly more progress. Examining the data, we would attribute their relatively strong performance to some combination of variables that predicted strong outcomes – organizational size or resources and readiness in terms of eagerness to adopt performance management and/or being further along on the benchmarks at the inception.

Finally, we have recommended that the BMA consider consolidating benchmarks 14 through 17. Devoting four out of 17 benchmarks to such a granular level regarding partnerships seems unnecessary and over-emphasizes its importance relative to the other benchmarks.

In short, given the time constraints, the benchmark assessment process suggests that all of the organizations demonstrated reasonably good progress in realigning themselves and building the systems and practices needed to support performance management.

Technical Assistance Work Plan

After completing the benchmark assessment tool, each organization met individually with Ellen Bass, the director of the Boston Capacity Tank. She played a pivotal role in conceiving of the project, raising funds for it, and managing it. After reviewing the self-assessment with the grantee and making any mutually agreed upon adjustments, Ellen and the grantee prepared a technical assistance work plan.

The first section of the plan described the intended outcomes, including a list of what the agency wanted to gain from their participation in the Institute. It also listed the 17 benchmarks from the assessment tool, organized by initial, intermediate and long-term outcomes, with scores denoting the performance management level on each benchmark at the outset of the project. The second section contains a detailed technical assistance activity plan, including the name of the capacity coach assigned to be the lead Institute technical assistance provider. The plan was presented in a seven-column table:

Capacity Building Activity	BCT Consultant	Agency Staff	Start Date	Deadline	Consultant TA Hours	Notes
----------------------------	----------------	--------------	------------	----------	---------------------	-------

The table contained a detailed and individualized workplan for each organization for the duration of the project. A third section was set aside for the final benchmark assessment at the end of the grant period to record the organization’s progress against each of the 17 benchmarks.

The plan also listed the materials the agency agreed to assemble for the consultant facilitating their theory of change workshop.

The Capacity Building Process

The Capacity Institute used four modalities to assist participating agencies to bring about organizational change, build knowledge about performance management and develop performance management practices. These were: 1) the theory of change workshop; 2) group training sessions; 3) individualized technical consulting; and 4) coaching. With the exception of the group training sessions, the assistance was customized for and delivered separately to each of the organizations.

1. Theory of Change Workshop

The theory of change workshop (sometimes referred to in the Institute's literature as a clinic) was a critical element in the program design. Each organization had a full-day workshop facilitated by David Hunter or Ellen Bass. Participating in the workshop were the organization's executive director and members of the program staff as well as one or more members of the agency's board. The workshop design was a modified version of a highly-structured 3-4 day process David Hunter developed to guide nonprofit organizations through a series of transformational steps needed to become highly-focused, outcomes-oriented and data-driven organizations, managed from top to bottom to achieve sustainable change for their program participants. Hunter has published articles about his theory of change workshop model; it is the signature product in his consulting practice and we are unaware of anyone else who facilitates a comparable workshop model.

Funding constraints led the Black Ministerial Alliance to prevail upon Hunter to shorten his usual 3-4 day process. The Capacity Institute's goal was to accomplish the most critical tasks in a one-day session. Funding was only sufficient to pay Hunter for 11 of the 15 theory of change workshops. Bass, who attended all of Hunter's presentations, used his PowerPoint slides and handouts, and even used some of his stories, during the workshops conducted for the remaining four organizations.

The workshop's overriding purpose was to convey a real understanding of what the organization would have to grapple with to become high performing. The universal set of issues to be addressed in making the transition included:

- Restating the organization's mission; one that would serve as a clarified statement of its value proposition – the specific social benefit it will deliver. The reason for starting with the value proposition is to get the organization to own what it purports to be producing.
- Translating the value proposition into a discrete set of goals. For Hunter, goals give focus to the mission statement. As important as defining the focus of an organization's activities, by their exclusion from these goals, the list effectively defines those things the organization will not do.
- Defining objectives establishes the measureable milestones toward achieving goals and facilitates the process of moving toward the concept of performance measurement.
- Specifying the target population is one more step in the larger exercise of creating focus. Just as goals narrow the scope of activities an organization will entertain, the target population definition serves as a restraint on the all too common impulse to respond to an ever expanding set of community needs. Achieving results requires this type discipline.

Defining the Bottom Line

- Codifying programmatic activities imposes further discipline on the organization by explicitly acknowledging the professional competencies needed and the intensity of services – the dosage, duration and frequency required to achieve results.
- Introducing some metrics that the organization can adopt to monitor activities, program utilization and outcomes.
- Identifying the specific indicators that will be used to track progress on outcomes.
- Describing the elements of the performance management system.

Compared to the 3-4 day process Hunter uses, the one-day theory of change session was more of a survey course, providing a sampling of steps towards creating a performance management culture.

The great challenge facing the Capacity Institute is its ambition to help organizations transition from an activities-centric approach to programs to one rooted in measurable outcomes. The starting point is the impacts the organization seeks to achieve rather than the activities it will engage in to secure those impacts. This approach involves a profound change in organizational culture, perspective, programmatic strategy and management discipline. As well intentioned as every agency was when they joined the Capacity Institute, organizational inertia nonetheless remained a powerful force that threatened to thwart the imperative to change.

An essential though unnamed strategy used by David Hunter and Ellen Bass in the theory of change workshop was to precipitate a crisis, one that would force a deep and honest dialogue about the fundamental, core mission of the agency and its strategy and methodology for advancing the organization's mission. From Hunter's and Bass' perspective, it was essential to expose and highlight fundamental inconsistencies, challenge deeply held assumptions, and confront the internal resistance that is present in every organization.

With only one day allotted for this work the facilitator had little time or opportunity to establishing a level of trust before challenging the organization. The results were mixed. Bass had already established a rapport with the organizations prior to the theory of change sessions she conducted. Her personality and style are less confrontational and aggressive than Hunter's. As a result her sessions were not as stormy as Hunter's. But both had to be disruptive to facilitate the change they sought to promote.

By most accounts, most of Hunter's sessions were quite uncomfortable. He pushed hard and had little patience for niceties. Despite his rough treatment however, most of the participants we interviewed were appreciative and felt they had accomplished a great deal with him. Some described it like a visit to the doctor, begrudgingly acknowledging that it was unpleasant but had a positive impact. Based on these interviews and observations of two of these sessions, it is our opinion that the challenging approach adopted by Hunter and Bass was in fact effective and essential to the model. Nonetheless, while the financial cost would have been substantially higher, it would have been helpful to devote more than one day to this pivotal exercise.

2. Group Training Sessions

Trainings were delivered in group settings or via webinar. These formats provided a cost-effective way to introduce all of the groups to foundational content knowledge. Together the training sessions amounted to a survey course on performance management for youth development organizations. The sessions followed the Capacity Institute's developmental sequence and related directly to one or more of the Capacity Institute's 17 benchmarks. While it was a coherent curriculum, as one participant explained, "I didn't understand how all the pieces related until the end of the project." The evaluators observed only one of the workshops but solicited and received feedback from participants and interviewed many of the presenters. The sessions included:

Positive Youth Development Framework – Since the intended impact of youth development organizations practicing performance management is to achieve positive youth outcomes and since the performance management framework hangs on a theory of change, this workshop laid the foundation for an evidence-based understanding of youth development. The content related to the first two benchmarks on the Institute's assessment tool. Delivered by staff of Health Resources in Action, the session's goals were to define youth development, articulate positive youth outcome, identify meaningful youth participation, and enhance programmatic strategies to build positive youth outcomes.

Logic Model Development – In preparation for each agency's preparation of program logic models, Bass and consultant Jin Min Lee conducted a training to introduce the terminology used for each portion of the logic model, to explain the structure and to familiarize the agencies with the idea of using theory of change. The third performance management benchmark focuses on both the existence and quality of logic models being used by the agencies participating in the Institute.

Outcome Measurement System Planning – Ellen Bass and Jin Min Lee also led this training session. Bass and Lee also served as the coaches. In that capacity they subsequently met with each organization to create an outcome measurement plan based on the logic models. The purpose of the training was to create a roadmap to help the agencies understand performance measurement operationally and preview the pieces of the plan each agency would need to develop over the coming months.

Communicating About Outcomes – While the Institute stressed performance management as a core practice that organizations needed in order to advance their mission, participants also understood that the funding environment continues to place greater emphasis on the ability of organizations to document their outcomes. RootCause delivered a training session about how to communicate outcomes information to external stakeholders. The trainers shared an agency report template as well as an example of a report for a youth development organization.

Client Relationship Management Database Software – Performance management is a data driven process. Few nonprofits have the type of sophisticated software to support data collection, report generation and data analysis. Yet it is essential to have this type of technology infrastructure in order to support the performance management system described in benchmark six. This session by

Defining the Bottom Line

Database Designs reviewed trends and options for monitoring programs and services with an eye to measuring success. In part, the session sought to introduce the organizations to the leading software options from which they could choose: Excel was included as one of the options to meet the needs of some of the smaller organizations participating in the Institute. Salesforce, an open environment that could be shaped to meet some data needs common to the participating organizations, and Efforts to Outcomes, a more costly proprietary solution were the other two products.

Introduction to Salesforce Software –Because Salesforce , provided several distinct advantages, including pricing, and the ease and speed of customization, Data Designs ran a follow-up training focused on that product. This training, like some of the others, provided an introduction to content that each would explore more deeply through individualized technical consultations. Each organization was eligible to receive \$4,000 worth of individual consulting with Data Designs. If a significant percentage of the organizations selected Salesforce, it would create economies of scale, and therefore cost savings, for training and programming special applications and templates. These advantages influenced the emphasis on Salesforce.

Selecting Measurement Tools – As part of United Way’s in-kind financial commitment to the Capacity Institute it hired the National Institute on Out-of-School Time (NIOST) to update information on United Way’s ToolFind website. Wendy Surr of NIOST then delivered a training to orient the agencies in how to find and evaluate valid and reliable data collection tools and to demonstrate the use of the site to search for evidence-based data collection tools. This training session prepared the agencies to fulfill the Capacity Institute’s performance management benchmarks four and five.

Human Resource Performance Management –A key to achieving internal alignment is a human resource performance appraisal system, documenting and connecting individual performance with organizational performance. Carlo Abrams of Blue Chip Consulting delivered the training that described an approach to job descriptions and performance appraisal in fulfillment of benchmark nine; a “human resource performance appraisal system, documenting and connecting individual performance with organizational performance, and giving feedback.”

Partnerships – Four of the performance management benchmarks, numbers fourteen through seventeen, focus on the importance of establishing and maintaining partnerships in the community to facilitate referrals more appropriately served by other providers; to connect youth in their programs to complementary services that meet a youth’s unique needs, and to liaise with participants’ school and/or employer. Part of the partnership relationship highlighted in the training was the need for the partners to have a common language and protocols.

Sustaining Change – As the Capacity Institute drew to a close, David Wright, BMA Executive Director, worked with Bass to deliver the final training session. The purpose was to focus the participating agencies on the transition from the BMA driving the performance management culture change to driving the performance management change themselves. They used an 8-step process

Defining the Bottom Line

described in John Kotter's *Leading Change* as the framework for the training. At the Institute's formal conclusion on September 30, 2012, none of the organizations were expected to have fully implemented their performance management system, nor mastered the 17 benchmarks. To succeed, each would have to persevere without the coach's prodding and deadline-setting. The more significant reason for building this Sustaining Change session into the Institute's design was the recognition that sustaining performance management practices depended on solidifying changes in the cultural values within organizations that the Institute worked to instill.

By and large, the participants found the training session either "helpful" or "very helpful." Occasionally one of the interviewees rated a session "somewhat helpful." We found no pronounced pattern that suggested any individual training disappointed the participants. The documentation reflects a good deal of thought about the selection of topics to buttress the overall structure and design of the capacity building program and the preparation done with the trainers prior to each session.

3. Individualized Technical Consulting

In general the training sessions introduced material needed to lay a foundation for a capacity building task each organization was about to undertake with the assistance of a technical consultant. For example, the logic model training preceded the series of working sessions the agencies had with a consultant advising them about and facilitating their efforts to develop program logic models. Similarly, the two training sessions about client relationship management software and Salesforce preceded the individual consulting sessions to customize and implement the software for each agency and train members of the staff in using it.

The feedback from the agencies was overwhelmingly positive about the quality and value of the consulting services. Several organizations complained about timeliness and follow-through with one consultant, although other agencies appeared to be satisfied or pleased with the same person's work. We shared the details of our findings about that individual consultant's performance with the BMA.

Without question, having technical specialists available to assist both with designing parts of the performance management system and assisting staff to learn new practices was indispensable. Like the coaching discussed below, the consulting sessions provided a structure that kept the organizations moving from step to step as they created their performance management systems.

4. Coaching

Observation and interviews confirmed the pivotal role coaching played in driving the change and learning processes for the agencies. Bass served as the coach. She acquired an intimate understanding of each organization and managed to establish very positive working relationships. Bass' unique qualifications for this role came through in our interactions and interviews with her over the course of this evaluation project. First, she is passionate about the power of performance management to produce needed impact in the community and has a missionary zeal to strengthen nonprofit practice. Second, she is a student of nonprofit management and organizational improvement and is continually learning and growing. Third, she is deeply respectful of the nonprofit sector and the staff and volunteers

Defining the Bottom Line

who are doing the work. Finally, she is able to push people to stay on track in a way that is supportive and encouraging, but consistent and firm.

References

- Blueprint Research & Design, Inc. (2002). *Marguerite Casey Foundation Organizational Capacity Assessment Tool*, Marguerite Casey Foundation: Seattle, WA.
- Hernandez, Georgiana and Visher, Mary G. (2001). *Creating a Culture of Inquiry. Changing Methods – and Minds – on the Use of Evaluation in Nonprofit Organizations*. James Irving Foundation: San Francisco
- Light, P. (2004). *Sustaining Nonprofit Performance: The Case for Capacity Building and the Evidence to Support It*. Brookings Institution Press: Washington, DC.
- McKinsey & Company, (2001). *Effective Capacity Building in Nonprofit Organizations*. Venture Philanthropy Partners: Reston, VA.
- Miller, Clara, (Summer 2008). “The Equity Capital Gap,” *Stanford Social Innovation Review*.
- Plantz., Greenway, M.T., & Hendricks, M. (2006). “*Outcome Measurement: Showing Results in the Nonprofit Sector*.” United Way of America: Alexandria, VA.
- Simister, N. with Smith, R. (2010). *Praxis Paper 23, Monitoring and Evaluating Capacity Building: Is it really that difficult?* International NGO Training and Research Centre: Oxford, UK.
- Wenger, E., McDermott, R., & Snyder, W.M. (2002). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business School Press: Cambridge, MA.