

The Unified Outcomes Project: Evaluation Capacity Building, Communities of Practice, and Evaluation Coaching

Jay Wade, M.A., Leanne Kallemeyn, Ph.D., and David Ensminger, Ph.D., Loyola University Chicago; Molly Baltman, M.A., Robert R. McCormick Foundation; and Tania Rempert, Ph.D., Planning, Implementation, and Evaluation Consulting Inc.

Keywords: Evaluation, evaluation capacity building, communities of practice, evaluation coaching, coaching

Key Points

- Increased accountability from foundations has created a culture in which nonprofits, with limited resources and a range of reporting protocols from multiple funders, struggle to meet data-reporting expectations. Responding to this, the Robert R. McCormick Foundation in partnership with the Chicago Tribune launched the Unified Outcomes Project, an 18-month evaluation capacity-building project.
- The project focused on increasing grantees' capacity to report outcome measures and utilize this evidence for program improvement, while streamlining the number of tools being used to collect data among cohort members. It utilized a model that emphasized communities of practice, evaluation coaching, and collaboration between the foundation and 29 grantees to affect evaluation outcomes across grantee contexts.
- This article highlights the project's background, activities, and outcomes, and its findings suggest that the majority of participating grantees benefited from their participation – in particular those that received evaluation coaching. This article also discusses obstacles encountered by the grantees and lessons learned.

Introduction

Advances in technological infrastructure for collecting, storing, managing, and accessing “big data” have furthered the use of data to understand and solve problems. Simultaneously, as foundations seek to maximize their investments,

a culture of increased accountability for distributed resources has been created, which translates into high expectations for reporting on outcomes. These circumstances require nonprofit organizations to develop some expertise in evaluation and data use.

The term evaluation capacity building (ECB) represents theoretical perspectives and practical approaches for addressing these circumstances. Integrating multiple definitions of ECB, Labin and colleagues defined it as “an intentional process to increase individual motivation, knowledge, and skills, and to enhance a group or organization's ability to conduct or use evaluation” (Labin, Duffy, Meyers, Wandersman, & Lesesne, 2012, p. 308). Based on a synthesis of empirical literature, they proposed an integrative model of ECB that is broadly composed of the need for ECB, ECB activities, and the results:

Collaboration between funders and projects may also be something to explore. Funders were not reported as being participants in the ECB efforts, but there was mention of their importance to the efforts. Adequate resources are needed not only to begin ECB efforts, but also to sustain them. If funders were included as target participants in the ECB efforts, it could increase their firsthand knowledge of ECB efforts and requirements, which, in turn, could affect expectations and funding cycles and reduce related resource and staff-turnover barriers. These hypotheses merit further exploration. (Labin, et al., 2012, p.324)

This article describes a case example of a collaborative ECB effort, the Unified Outcomes Project, an initiative sponsored by the Robert R. McCormick Foundation among 29 social service agencies receiving funding through the Chicago Tribune Charities, a McCormick Foundation fund. The project's aim was to increase collaboration between the funder and their grantees and mutual understanding about funder needs and grantee realities. This article focuses on two specific mechanisms that facilitated these outcomes: communities of practice (CP) and communities of practice with coaching (CPC). Multiple ECB models (Preskill & Boyle, 2008; Labin, et al., 2012) note that a combination of ECB strategies, including coaching and CP, are associated with higher levels of organizational outcomes. In comparison to previous case examples (Arnold, 2006; Stevenson, Florin, Mills, & Andrade, 2002; Taut, 2007; Ensminger, Kallemeyn, Rempert, Wade, & Polanin, 2015), the Unified Outcomes Project focuses on the mechanisms of CP and CPC to highlight a unique approach to ECB that could potentially be used across various foundation contexts.

Background and Need

The behavioral health and prevention field is complex and without a unified set of outcomes embraced by all professionals in the area, as exists in fields such as workforce development (e.g., percentage of clients placed, salary, job retention) and homelessness (e.g., percentage of clients maintaining permanent housing). Although measurement tools exist to assess the impact of behavioral health and prevention services (e.g., decrease in trauma, increase in functioning, increase in parenting skills), it was unclear to the foundation which of these tools was effective in measuring the impact of treatment and capturing information in a culturally appropriate manner. Also, through discussions during site visits, grantees running similar programs expressed conflicting views about using specific evidence-based tools.

To address these issues, the foundation began to consider ways to improve evaluation within the child abuse prevention and treatment funding area. Program staff wanted to be able to compare program outcomes using uniform evaluation

Program staff wanted to be able to compare program outcomes using uniform evaluation tools and to use that data to make funding, policy, and program recommendations, but they were at a loss as to how to do so in a way that honored the grantees' knowledge and experience. A newly hired director of evaluation and learning advised staff to strongly encourage evaluation and include grantees as partners in the planning and implementation processes as a cohort group.

tools and to use that data to make funding, policy, and program recommendations, but they were at a loss as to how to do so in a way that honored the grantees' knowledge and experience. A newly hired director of evaluation and learning advised staff to strongly encourage evaluation and include grantees as partners in the planning and implementation processes as a cohort group.

With this direction, foundation staff spoke individually with grantees to introduce the ideas of unifying outcomes, creating an evaluation learning community, and providing capacity-building support. Although grantees differed in their initial enthusiasm for such a project, foundation personnel felt that there were enough grantees interested to proceed. Thus, the Unified Outcomes Project was initiated with the hope that, with transparency and inclusiveness, it could:

An evaluation coach works with stakeholders to facilitate the development of the attitudes, beliefs, and values associated with conducting evaluations, along with knowledge and skills. Evaluation coaching promotes these dispositions through different types of coaching and the facilitation of various learning processes, such as relating, questioning, listening, dialogue, reflecting, and clarifying values, beliefs, assumptions, and knowledge

1. Benefit grantees by building their evaluation capacity.
2. Improve existing programs through use of evaluations and data.
3. Improve the foundation's funding decisions by creating a unified set of reporting tools across grantees in the child abuse prevention and treatment funding area for grantmaking decisions.
4. Ultimately help children and families.
5. The foundation hired an evaluation coach to facilitate the project's progress and build grantee evaluation capacity.

The decision to hire an evaluation coach was intentional, as the goal of the foundation was to

support the programs in building evaluation capacity for the purpose of organizational learning. To promote evaluation capacity, organizations often need to shift toward a learning framework (Preskill & Boyle, 2008), which requires genuine dialogue, developing trust, open-mindedness, and promoting participation (Preskill, Zuckerman, & Matthews, 2003; Torres & Preskill, 2001). The competencies needed to support an organization's shift extend beyond the technical knowledge of and skills for conducting external evaluations, and requires competencies associated with coaching (Ensminger, et al., 2015). An evaluation coach works with stakeholders to facilitate the development of the attitudes, beliefs, and values associated with conducting evaluations, along with knowledge and skills. Evaluation coaching promotes these dispositions through different types of coaching and the facilitation of various learning processes, such as relating, questioning, listening, dialogue, reflecting, and clarifying values, beliefs, assumptions, and knowledge (Ensminger, et al., 2015; Griffiths & Campbell, 2009; Torres & Preskill, 2001). With an evaluation coach on board, the project began in earnest to:

1. Agree on a set of outcome data to be collected across all grantees.
2. Create CP in conjunction with evaluation coaching.
3. Build evaluation capacity with participating grantees.
4. Promote cross-organizational learning.

Role of the Evaluation Coach

The purpose of the evaluation coach was to facilitate each cohort's CP meetings, synthesize and systematize cohort reporting tools, and lend additional support via one-on-one coaching to grantees that requested it. One-on-one coaching sessions provided support to the grantees on administering the tools, collecting and analyzing data, and reporting findings in a comprehensive, meaningful manner. The coaching was dynamic; the coach adjusted the type of evaluation assistance to the level of a grantee's existing evaluation

capacity. In most circumstances, this meant the one-on-one evaluation coaching expanded beyond the specific tools and outcomes identified in CP meetings to the particular evaluation needs of each organization, independent of the project's goals.

The evaluation coach met the grantees in person at their offices. Being on-site was an important component, helping the evaluation coach experience how explicit and implicit protocols were implemented in practice. Having a better understanding of how and why processes did or did not work for a specific organization enabled the coach to tailor her coaching for the organization to support its individual ECB goals. With some grantees, the coach worked on the most basic level with staff to define a theory of change and develop logic models. Other grantees had a department devoted to evaluation, and the coach worked with clinical staff's use of evaluation information to improve service quality and evaluation buy-in. The in-person, needs-oriented approach of the coaching sessions helped build coach-organization rapport and developed a "personal factor," which promotes better evaluation outcomes and use (Patton, 2008). Although the individual agencies each worked with the evaluation coach on specific activities, outputs, and outcomes, the goal of the one-on-one coaching was to improve the quality and efficiency of evaluation practices by helping grantees to develop their own internal capacity for quality program evaluation.

Unified Outcomes Project Activities

Phase One: Unifying Outcomes

Foundation personnel and the evaluation coach scheduled a initial meeting to introduce the ECB project, inviting all 29 grantees. At this meeting, they gathered input from the grantees on the frustrations and benefits of evaluation, data collection, and reporting. These discussions revealed that grantees were using a multitude of tools and felt burdened by the work required to implement them and report findings. It was agreed that tools should focus on three specific areas: improvements in parenting, increases in children's behavioral functioning, and decreases in child trauma symptoms. Based on these distinctions, the foun-

They gathered input from the grantees on the frustrations and benefits of evaluation, data collection, and reporting. These discussions revealed that grantees were using a multitude of tools and felt burdened by the work required to implement them and report findings.

dation and the evaluation coach convened a second meeting, dividing the grantees into three cohorts representing their program services: positive parenting, child trauma, and domestic violence. These cohorts became communities of practice to address these service areas.

The CP meetings in this phase of the project consisted of two half-day sessions where each cohort convened at the foundation with McCormick personnel and the evaluation coach. At the first meeting, grantees discussed in more detail how evaluation practices were being used in their programs, including their favored assessment tools and data they were required to report to public and private funders. Grantees reported a total of 37 tools to the foundation. Participants discussed each of the assessment tools' strengths and weaknesses, focusing on the length, developmental appropriateness, and language (i.e., strengths-based language versus deficit language) of the tools as well as the alignment of each tool to program outcomes and the grant application.

After these discussions, foundation staff in collaboration with the evaluation coach sent an electronic survey to all grantees asking about their preferred client-assessment tools, what they were required to collect and report by other funders, best practices they wanted to represent with mea-

All grantees were able to identify a total of six common tools they were willing to use – one to three tools per program area. The foundation agreed to require at least one of those six tools, so every organization was able to use a tool that was either its first choice or one it identified as willing to use. None of the grantees would have to report on tools that were their last choice or that they would use only if required by the funder.

surement tools, and program-level outcome questions. The results showed wide agreement among the grantees. Drawing on previous CP discussions, all grantees were able to identify a total of six common tools they were willing to use – one to three tools per program area. The foundation agreed to require at least one of those six tools, so every organization was able to use a tool that was either its first choice or one it identified as willing to use. None of the grantees would have to report on tools that were their last choice or that they would use only if required by the funder.

At the second CP meeting for each cohort, the list of common tools was revealed, and the grantees were pleased that they would not be required to use a tool that did not fit with their program. The evaluation coach then led each cohort through a detailed discussion and training on implementing the common assessment tools, including developing a protocol all grantees would follow on the

timing of pre- and post-tests, client eligibility for testing, and data collection. The coach worked individually with grantees at their request to develop protocols that fit each organization's culture. In addition, four grantee staff members who were the most knowledgeable in their fields and had already integrated evaluative thinking into their agencies were asked to serve on an advisory group that would give input into the surveys, professional-development workshops, and materials developed as part of the initiative.

Phase Two: Evaluation Capacity Building

During the second phase the evaluation coach facilitated six half-day, in-person CP meetings, which served as professional development for grantees on evaluation topics identified by the cohorts. Each cohort had specific questions and concerns related to evaluation practices and tool implementation. Agendas for cohort meetings were based on these concerns and requests – grantees were helping to set the agenda. The coach also developed automated reporting dashboards for the tools each cohort selected.

Grantees were also offered coaching support at three levels of intensity. Level one, the lowest intensity, entailed only participation in CP meetings with the cohort throughout the year. At level two, grantees received both the CP meetings and the opportunity to work with the evaluation coach individually during the year to assist with the implementation of the new tool or tools. Level three provided the components in the other two levels as well as support on a range of evaluation topics beyond the scope of implementing the new tools, such as logic modeling and using data for program improvement. The goal of level three was to create an evaluation culture with grantees and further build their evaluation capacity. Not all agencies needed or wanted the third level of coaching, and each agency was encouraged to choose the level that seemed most appropriate for their organization. In practice, grantees that initially chose level-two support ended up engaging the coach and process at the same intensity as the level three grantees. As the evaluation coach began meeting with level-two grantees, the coaching

naturally began to extend beyond the implementation of the tools as each grantee expressed other evaluation needs. At CP meetings, grantees heard about the benefits of the coaching from other grantees and began to engage the coach more frequently. Thus, in practice, there were two types of grantees, those who received level-one (CP) support and those who received level-three (CPC) support. Of the 29 grantees, 14 chose CPC and 15 chose CP.

Phase Three: Benchmarking and Practice

With evaluation coaching and capacity building ongoing, the project's focus shifted to benchmarking grantee practices based on grantee feedback and input. Convening the cohorts to discuss the grant application, the foundation and the evaluation coach revamped the application based on their suggestions. The rubric for assessing the grant application was also shared with grantees to gather their input and share their suggestions for the program officers to more effectively rate applications. Once the foundation received feedback from each cohort on the application and rubric, the advisory group reviewed the final draft and identified sections of the rubric to be weighted for importance when assessing a program. Foundation personnel used the updated application and new rubric during the June 2015 funding cycle. The rubric captured program indicators beyond assessment (i.e., qualitative data), allowing foundation staff to compare agencies in a more holistic manner.

Methods

The research team used case study methodology (Stake, 1995; Yin, 2014) to study the Unified Outcomes Project. Interviews of grantee participants, observations of CP and CPC sessions, and the Evaluation Capacity Assessment Inventory (Taylor-Ritzler, Suarez-Balcazar, Garcia-Iriarte, Henry & Balcazar, 2013) were used to gather evidence of outcomes and obstacles to ECB. Twelve interview participants were selected via a collaborative process among the researchers, foundation program managers, and evaluation coach. The goal was to sample across varying levels of project participation (i.e., CP and CPC), evaluation capacity, and the size of the program budgets.

In practice, grantees that initially chose level-two support ended up engaging the coach and process at the same intensity as the level three grantees. As the evaluation coach began meeting with level-two grantees, the coaching naturally began to extend beyond the implementation of the tools as each grantee expressed other evaluation needs. At CP meetings, grantees heard about the benefits of the coaching from other grantees and began to engage the coach more frequently. Thus, in practice, there were two types of grantees, those who received level-one (CP) support and those who received level-three (CPC) support.

The research team, coach, and foundation staff convened to assess each organization's evaluation capacity. This was determined by three criteria: the Evaluation Capacity Assessment Inventory (ECAI), which was administered to each grantee in project at the beginning of Phase Two (Taylor-Ritzler, et al., 2013); how thorough and timely each grantee reported its program evaluations to

TABLE 1 Interview Sampling

Grantees Sampled for Interviews as Described by Evaluation Capacity and Program Budget			
	High Evaluation Capacity	Medium Evaluation Capacity	Low Evaluation Capacity
High Budget <\$400,000	Grantee No. 5 Grantee No.8	Grantee No. 12 Grantee No. 10	Grantee No. 7 Grantee No. 4
Low Budget >\$400,000	Grantee No. 1 Grantee No. 6	Grantee No. 9 Grantee No. 3	Grantee No. 11 Grantee No. 2

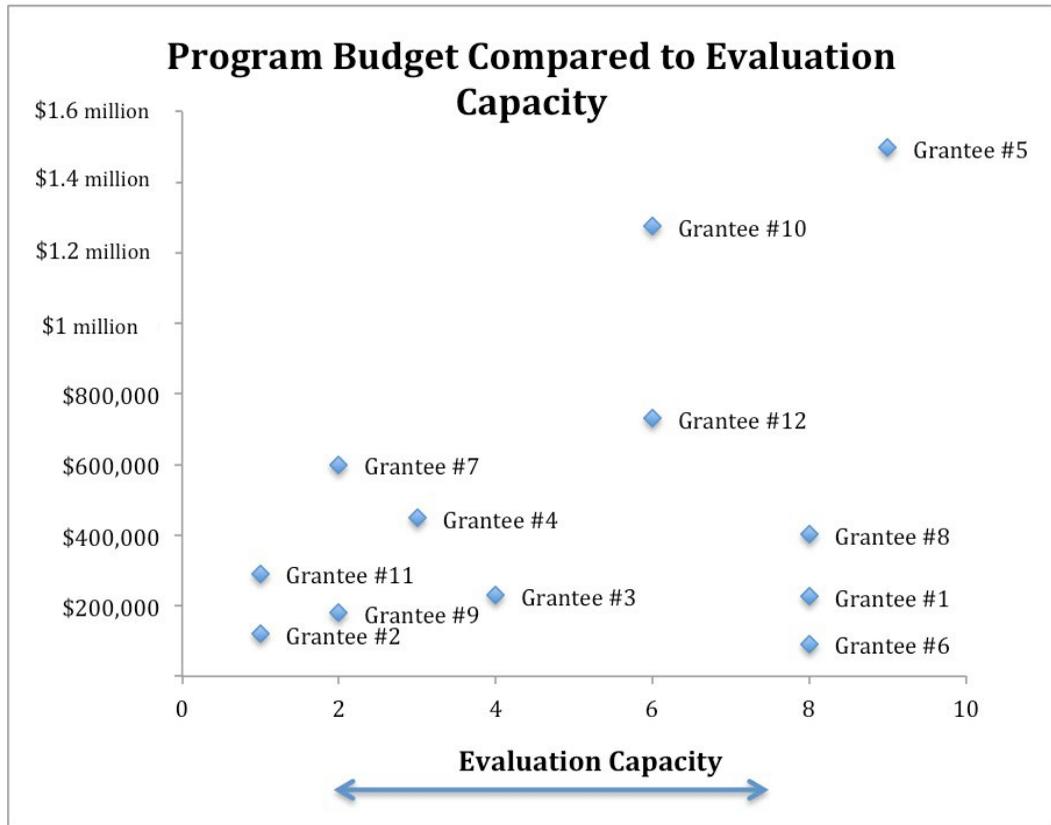
the foundation; and grantee leadership and attitudes toward evaluation as judged by project participation in the cohort meetings and one-on-one coaching sessions. Using these three criteria, grantees were categorized into high, medium, and low evaluation-capacity levels. A high-capacity grantee typically had an internal evaluator or evaluation department that facilitated the development of logic models and collection and analysis of outcome measures, and routinely and with ease submitted complete reports to the foundation. A medium-capacity organization typically employed staff whose job descriptions included evaluation, made some use of logic models and outcome measures, and were generally able to complete reports for the foundation, although systematic processes for doing so were not in place. A low-capacity grantee had no staff dedicated to evaluation and had difficulty providing complete and timely reports. Grantees were also categorized by their program budgets: The median budget for grantees involved in the project was \$400,000; those below that were categorized as “low budget” and those above the median were categorized as “high budget.”

The research team selected 12 grantees across capacity levels for interviews, including six CP grantees, six CPC grantees, and grantees that ranged between high and low budget. (See Table 1.) The goal was to have one CP and one CPC grantee of both high, medium, and low evaluation capacity at the start of the project as well as high and low budget. While this ideal was not realized (there was no CP grantee categorized with medium evaluation capacity and high budget), care was taken to make sure that this goal was maximized. (See Figure 1.)

A hermeneutical approach (Kvale & Brinkmann, 2009) was utilized during the analysis. This approach is not a step-by-step process, but rather involves adhering to general principles of interpretation. Key principles include a continuous back-and-forth between parts and the whole to make meaning, such as experiences of one grantee in the relation to the entire sample; a goal of reaching inner unity in the findings; awareness that the researchers influence the interpretations; and the importance of the interpretations promoting innovation and new directions. During this process, the research team applied ECB frameworks (Preskill & Boyle, 2008; Labin, et al., 2012) and allowed for emergent themes. Frequent meetings were held to gain consensus among the research team, evaluation coach, foundation staff, and selected participants.

The ECAI was administered to all grantees six months into the project and a year later, at its conclusion (Taylor-Ritzler, et al., 2013). Scores for nearly all grantees decreased from pre-test to post-test, which was explained well by Grantee No. 3: “I think when it comes to evaluation, partly it’s challenging because I don’t know what I don’t know, right?” This demonstrates response-shift bias (Howard & Dailey, 1979), a phenomenon in which participants’ pre-test responses are often higher estimates than their actual ability because they have not yet been exposed to an intervention. Anticipating response-shift bias, a single “perceived change” item was added at the conclusion of each construct at post-test so participants could gauge their own growth over the course of the year (e.g., “Based on my participation in the McCormick project, I believe mainstreaming has increased.”). Due to response-shift bias and the triangulation of the interviews and observations

FIGURE 1 Evaluation Capacity vs. Program Budget



with the single perceived-change item, results discussed in this article are based on the scores of these adapted items. The statistical authority of the ECAI results should be understood in light of a low number of grantee responses ($n = 33$ individual responses; some grantees had multiple staff respondents). Thus, ECAI results are discussed only in relation to the interview data.

Findings and Reflections on the Unified Outcomes Project

Models of ECB can serve as a lens for understanding grantees' perspectives on their experiences with the Unified Outcomes Project. Strategies from Preskill and Boyle's (2008) ECB model that were most evident in this project included CP and coaching, although we considered all ECB strategies described in the model. Grantees' perceived outcomes also aligned with constructs in Labin, et al.'s (2012) ECB model, as well as the ECAI (Tay-

lor-Ritzler, et al., 2013). We organized our findings based on the salient changes in: (1) processes, policies, and practices for evaluation use; (2) learning climate; (3) resources; (4) mainstreaming; and (5) awareness of and motivation to use evaluation.

Within the description of these outcomes, we distinguished the shared and differential impact of CP and CPC. First, CP provided grantees and the foundation an opportunity to reflect critically on data-collection tools and processes. Second, CP facilitated a learning climate within the grantee organizations, although not consistently across grantees. Third, grantees viewed the evaluation coach as a key resource. Fourth, two grantees reported mainstreaming evaluation practices within their respective organizations, which facilitated its use. Although grantees were still integrating these practices and faced obstacles to mainstreaming during data collection, those that participated

TABLE 2 Grantees' Perceived Change of ECB constructs after 18 months on adapted ECAI items

Grantees' Perceived Change of ECB Constructs After 18 Months on Adapted ECAI Items (n=33)			
Construct	Level	Item Scores	Difference
Awareness of Evaluation	CPC	3.6 (0.69)	+0.83*
	CP	2.77 (1.17)	
Motivation	CPC	3.5 (0.7)	+0.25
	CP	3.25 (0.87)	
Competence	CPC	3.44 (0.73)	+0.34
	CP	3.1 (0.94)	
Leadership	CPC	3.13 (0.84)	+0.13
	CP	3.0 (0.67)	
Learning Climate	CPC	3.5 (0.76)	+0.6
	CP	2.9 (0.74)	
Resources	CPC	3.38 (0.74)	+1.18**
	CP	2.2 (0.83)	
Mainstreaming	CPC	3.22 (0.83)	+0.78
	CP	2.44 (0.73)	
Evaluation Use	CPC	3.11 (0.78)	+0.81
	CP	2.3 (0.95)	

*Indicates a statistically significant result at the p <0.05 level
 **Indicates a statistically significant result at the p <0.01 level
 "Strongly disagree" = 1, "somewhat disagree" = 2, "somewhat agree" = 3, "strongly agree" = 4

in CPC particularly benefited in this area. Finally, individuals reported some benefits to their awareness of and motivation to use evaluation. Less impact in these areas may also be attributed to these grantees and their representatives entering the project with some general competence in evaluation and positive attitudes toward evaluation. Similarly, no grantees discussed changes in leadership. The minimal discussion of leadership might be an artifact of whom we interviewed, since the participants selected for the project were leaders in their organizations.

Based on the in-depth interviews, 11 of the 12 grantees described at least one outcome from the project, and some grantees described as many as five. (See Figure 2.) CPC grantees reported more outcomes than did CP grantees. Results from the perceived-change items on the ECAI triangulated with the findings from the interviews. (See Table 2.) Overall, CPC grantees reported more growth than did CP grantees. Although grantees at all levels of evaluation capacity reported outcomes, the

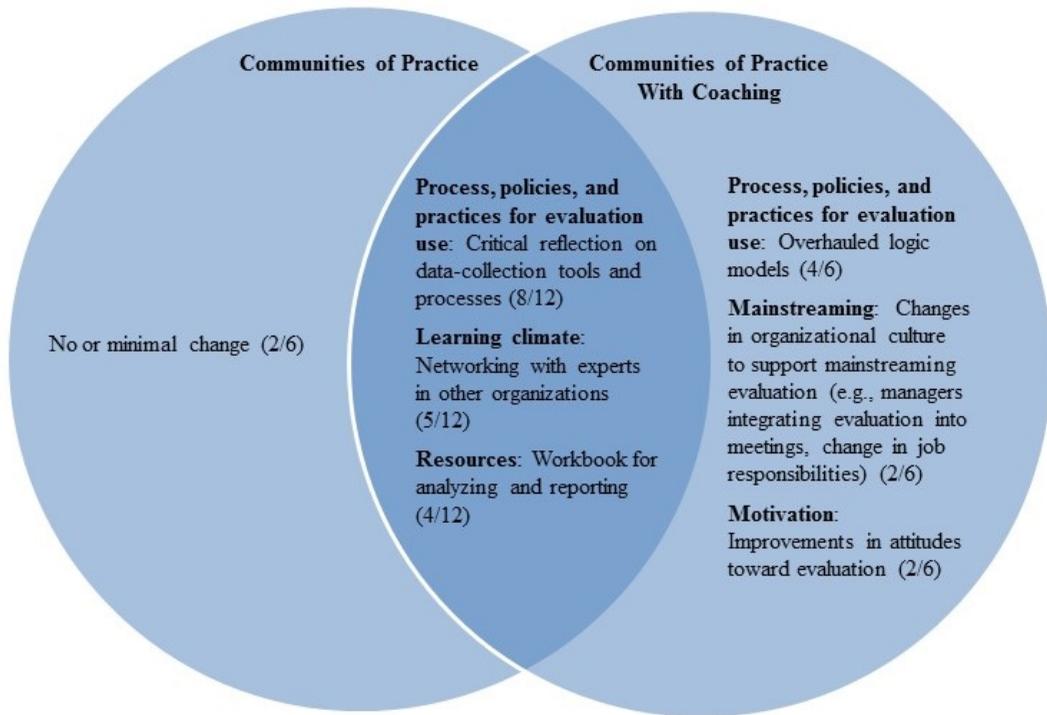
project seemed to have more impact on grantees with medium capacity than on the grantees with high and low evaluation capacity. Across grantees and reported outcomes, there were 12 out of 24 possible instances of outcomes for grantees with medium evaluation capacity, whereas grantees with low and high capacity had less – five and seven out of 24, respectively.

Critical Reflections on Data-Collection Tools and Processes

Participation in CP resulted in shared outcomes across grantees. (See Figure 2.) Grantees most commonly discussed the results of critically reflecting on their outcome tools, eliminating unnecessary tools, adopting more appropriate tools, and developing processes to utilize them. Grantee No. 12, who had a high budget and medium evaluation capacity and who received evaluation coaching, described the experience:

What we found was that we were using a lot more evaluation tools than a lot of other places. ... It really

FIGURE 2 Venn Diagram of Outcomes



made us look at why we were using everything that we were using. Then the one-on-one with [the evaluation coach] was really helpful to really look at what we were using with the results of those evaluations. So instead of looking at a snapshot each year, it was more, how can we use this tool right now to turn that around and help our participants throughout the year?

Later in the interview, the same participant reported, “We stopped doing that 160 questions; we’ve started new tools; we’ve had more conversations around how we do these tools, what we do with the data – so we’ve made so many changes already.”

These accounts demonstrate how CP facilitated a critical reflection on evaluation processes and practices that facilitated evaluation use. Grantees that maintained their tools also described the process of critical reflection as valuable, particularly as it helped them evaluate the extent to which the

tools they were utilizing were culturally relevant with their clients.

In addition, the majority of the grantees that received coaching described how they “overhauled” their logic models with the assistance of the coach. Grantees reported revisiting, updating, and creating logic models for additional elements of their programming and developing deeper understandings of the elements in the logic models.

Despite the changes made within their organizations, almost all – 11 – grantees still identified a lack of systemized process and practices as a limiting factor in their ability to fully develop their evaluation capacity. Participants noted either absent or limited mainstreamed practices and processes for data collection, analysis, and reporting. Grantee No. 10 reported a

... lack of systems, lack of protocols, and lack of administrative structure to actually facilitate the timely

This sense of community among grantees developed as the process progressed. The CP meetings had an affective role at the beginning stages, but as reporting started to take place the grantees shared with one another tangible ideas and efforts that included a successful dashboard tool, a system for reporting, and ways to promote clinical buy-in of evaluation. The grantees' struggles became, along with organizational mission, a bond that led to dialogue and ultimately served as a valuable tool in the development of evaluation protocols and procedures.

collection and [reporting] of data. So an example might be, we have quarterly reports due for our [Department of Children and Family Services] contracts and reports. ... So, part of the scramble – oh, we need that data and there's not one person who tracks all of it, or it's requested from the clinicians so we'll have to go back to the clinician to get that information in their caseload, and their score. So it's kind of a lack of administrative structure and protocol.

Even though grantees changed evaluation practices and processes through streamlining tools, they continued to encounter obstacles in this area.

Learning Climate

Half of the grantees explicitly attributed their changes in practice to the “learning” and “dialogue” that took place during these CP meetings. Consistent with the interview findings, the adapted ECAI item related to learning climate did not show a statistical difference between CP and CPC grantees. (See Table 2.) Grantees shared in interviews that substantive learning occurred when experts within the community shared their knowledge on particular evaluation tools. Grantee No. 12 described a participant from a high-capacity organization as “awesome”:

She is like my evaluation guru. I would love to talk to her more on an ongoing basis. ... I've just gone to her training on that [tool] and, based on her training, we decided to do that one as well.

In addition to utilizing experts within the diverse group of grantees, CP provided a forum to negotiate collectively with the program officer. Grantee No. 4 remarked,

It's always nice to come together with the other [domestic violence] agencies, because you're not the only voice saying the same things. So it was nice to have other people echoing, “there's the data we can collect; here's what we can do; I can't really give you that, but I can give it to you this way.” So it was nice to have the shared voice and kind of get an idea of what other agencies do.

Common issues among the grantees, such as data use and interpretation, led to discussion when the grantees convened at the foundation. These CP meetings facilitated a dialogue among participants about their own practices, successes, and failures. The secondary benefit of this dialogue was that the grantees felt validated and “heard” by the foundation. Grantee No. 10 commented,

We are all doing a lot with a little. ... We all have our struggles and challenges ... but we ... have a common mission for our target population. And it shows,, that it's okay to voice your challenges with your funders.

This sense of community among grantees developed as the process progressed. The CP meetings had an affective role at the beginning stages, but as reporting started to take place the grantees shared with one another tangible ideas and efforts that included a successful dashboard tool, a system for reporting, and ways to promote clinical buy-in of evaluation. The grantees' struggles became, along with organizational mission, a bond that led to dialogue and ultimately served as a valuable tool in the development of evaluation protocols and procedures.

Although grantees experienced a learning climate at CP meetings, individual grantees reported difficulty developing it within their own organizations. Six grantees described climate limitations that were interfering with ECB, such as staff buy-in of evaluation and viewing evaluation as a way to improve programming. Said Grantee No. 7,

I think the [learning climate is] emerging. I think there's still some sense of trying to understand how it's incorporated By and large, leadership gets the difference. But I think there's still the challenge is it compliant, and quality assurance, and if that's different than evaluation.

Thus, grantees had difficulty transferring the learning climate from CP into their own organizations.

Resources

Once the common evaluation tools were adopted, the evaluation coach developed resources – such as an Excel workbook – that allowed grantees to enter data and generate a report for the foundation and for their internal use. During the CP meetings, the coach trained grantees to use these workbooks. Grantee No. 12 observed, “I often have to say in my reports, ‘data not yet available,’ and I’d really like to be able to report ongoing about it, which is awesome now that we have the workbook.” Using the evaluation tools this way required many grantees to shift from using the tools clinically to aggregating them for program evaluation. Grantee No. 3 said,

Grantees who received coaching viewed it as an opportunity to improve their organization’s evaluation procedures according to its specific needs. Grantees who did not participate in the coaching could have benefited from it, but saw the process as undefined.

I think we have all felt pretty comfortable, like, “I gathered this data and here's what it meant to me clinically with this family. ...But I ... pull back and think ..., “How can we use data in the aggregate to make program decisions?”

The resources provided to the programs through the coaching and spreadsheets for analyzing their tools were supporting the use of data for program improvements. Results of the adapted ECAI items showed that CPC grantees reported higher growth than CP grantees for resources. (See Table 2.) These findings were statistically significant.

Questions about data were frequently discussed with the coach. Although the struggles with data use and interpretation were shared across grantees, each grantee's data needs depended on its context and mission. Grantee No. 11 said the coach “actually [came] out to us, to actually look at our particular challenges. ... She was able to ... view what our specific challenges were and then [ask] us what would be the most helpful currently, right now.” All grantees benefited from the nonlinear, needs-oriented coaching approach. As Grantee No. 8 described it, “If we were going to use this consulting service, we wanted it to meet

Eight out of the 12 grantees cited lack of time as an obstacle to developing evaluation capacity. Grantees recognized that for evaluation to benefit their services, they needed time to engage in activities beyond simple outcome measuring.

us where we were. And [the coach] was very open to that.” In fact, this appeared to be a key difference between grantees who chose the coaching and those who did not. Grantees who received coaching viewed it as an opportunity to improve their organization’s evaluation procedures according to its specific needs. Grantees who did not participate in the coaching could have benefited from it, but saw the process as undefined.

Grantee No. 4, who did not participate in coaching, explained,

We are trying to come up with a better way of tracking service-plan outcomes ... that's the kind of coaching I would more prefer, rather than around formal assessment tools.

This type of coaching was available, but some grantees did not understand that it extended beyond the outcome measures reported to the foundation.

These coaching outcomes were not simply a result of technical expertise, but also depended upon the coach’s demeanor and attitude. Grantee No. 11 reported,

There’s no pomp and circumstance with her, OK? It’s direct information, clearly stated in layman’s

terms. ... This was not some university professor come to lecture to us. This was a person who really rolled up her sleeves and worked with us.

The progress of these grantees was due not only to technical experience, but to the disposition of the coach, who helped to clarify and engage difficult material.

Even though the grantees viewed the coach and the tools as resources, they also cited a lack of resources as a continuing limitation. Nine of the 12 grantees mentioned personnel as a barrier to their capacity-building efforts. Several grantees said evaluation was not a central part of any one person’s job responsibilities and that their organizations did not employ personnel explicitly for evaluation work. “I would not say that my main job description involves [evaluation],” said Grantee No. 2. “It’s a part of my job description, ... a part of what I have to do with every client, but it’s not my main job. So we don’t have ... an evaluation person.”

Six out of the 12 grantees reported technology as a barrier to building evaluation capacity. Most grantees did not have software to manage and analyze data, which hindered the organization’s ability to do ongoing evaluation. Said Grantee No. 12, “We do so much evaluation, but it’s kind of scattered. ... If we could ever have more money to have a better database so that everything could go into one area, we could pull data from one area.”

Eight out of the 12 grantees cited lack of time as an obstacle to developing evaluation capacity. Grantees recognized that for evaluation to benefit their services, they needed time to engage in activities beyond simple outcome measuring. “There needs to be much more time spent on data management, data collection, data cleaning, and linking outcome data to other data and demographics,” Grantee No. 1 said. Although the grantees who participated in coaching viewed the evaluation coach and the tools she provided as resources for facilitating ECB, the grantees still lacked resources for evaluation.

Mainstreaming

Only two grantees described outcomes related to mainstreaming evaluation, but such change was substantial and important. Grantee No. 12, for example, changed job responsibilities to incorporate evaluation:

Now I am the person who collects all the evaluation and does the data entry into the workbook that [the coach] created for us. Then I work with the managers to help them determine how they will use that evaluation in their programs.

The same grantee integrated discussion of evaluation into organizational routines of meetings with managers that facilitated the ongoing use of evaluation:

I've added evaluation to every meeting. ... I've sat down with my managers recently and looked at what questions they're asking compared to what other managers' programs are asking. Because it just seemed very inconsistent. ... It just didn't make sense that only Healthy Families would be asking certain questions, when our doula participants should be getting asked the same things. ... I think that way, all of our managers have had more buy-in about committing to getting that information and why we're asking for it.

This account demonstrates how the foundation was able to model an inquiry and capacity-building process that a director replicated within the organization.

Although only two grantees discussed mainstreaming evaluation during the interview, grantees who received coaching “somewhat agreed,” on average, that their participation was related to changes in mainstreaming and evaluation use; grantees that participated only in CP “somewhat disagreed” on average. (See Table 2.) These subscales approached significance: mainstreaming ($p = .051$) and evaluation use ($p = .059$).

Awareness of Evaluation and Motivation to Use Evaluation

CPC grantees scored higher than CP grantees on the adapted ECAI item for awareness of evalu-

Grantees reported limited evaluation knowledge and skills among other personnel in their organizations, which restricted the development of evaluation capacity.

ation. (See Table 2.) This difference was statistically significant. In contrast to the ECAI findings, grantees did not extensively discuss changes in their awareness of evaluation during interviews. If interviews had also been conducted with grantees at the beginning of the project, these changes may have been evident. Two grantees, both of whom participated in CPC, did describe improvements in their attitudes or motivations for using evaluation, although related items on the adapted ECAI did not have differences between grantees in CP and CPC. In comparison to other areas discussed, motivation to use evaluation and competence in evaluation were also strengths of the grantees at the outset of the project.

In interviews, grantees reported limited evaluation knowledge and skills among other personnel in their organizations, which restricted the development of evaluation capacity. Eight of the 12 grantees reported that service-delivery staff's limited understanding of evaluation was an obstacle that needed to be addressed, and seven of the 12 grantees reported that limited knowledge and skills of those in director-level positions hampered ECB. Grantees needed to train the front-line staff on the use of the outcome-measures tools and the role the tools played in program monitoring. Grantee No. 4 said, “Our barrier is more [that] the counselors aren't seeing quite as much the utility of using the formalized tool.” As in other areas, grantees were wrestling with ECB within their own organizations, even though individual participants in the project demonstrated some evidence of ECB.

Program officers have seen big differences in the ways grantees have changed their internal culture of evaluation with direct-service staff, reduced the time agency staff spend on collecting and reporting data, and improved their ability to show impact through data using evidence-based evaluation tools.

In summary, two strategies – CP and coaching – facilitated outcomes related to critical reflection on data-collection tools and processes, learning climate, resources, mainstreaming, awareness of evaluation, and motivation to use evaluation. CP helped link grantees with similar missions and create a network of learning and reflection on evaluation practices. Furthermore, the convening of grantees and funders helped build a more nuanced understanding of common struggles and issues grantees often faced. Some grantees who chose to participate in CPC reported additional benefits, including overhauling and creating logic models, additional resources for ECB, mainstreaming evaluation practices, and a greater awareness and motivation to use evaluation.

Foundation Perspective

Foundation personnel who were deeply involved in this process were asked to write and reflect on the project and its outcomes. Presented here is a reflection from Molly Baltman, the McCormick Foundation's assistant director of grantmaking for community programs and this article's second author, on the outcomes, implications, and future directions of this project:

Throughout the past two years, a new culture of learning and peer support has developed between

grantees and the foundation, one that has ultimately led to program improvement. Grantees who were initially resistant to this process are now open to exploring ways to improve program services through evaluation practices. This shift occurred largely due to observations of other grantees having successes and embracing evaluation for learning and quality improvement. Program officers have seen big differences in the ways grantees have changed their internal culture of evaluation with direct-service staff, reduced the time agency staff spend on collecting and reporting data, and improved their ability to show impact through data using evidence-based evaluation tools.

The original plans of the project did not include the intensity of evaluation coaching that the foundation ended up supporting. It was anticipated that through CP meetings, grantees would learn and share evaluation approaches and practices that would build on the high-quality services already provided. The foundation had not anticipated the level of hands-on technical assistance grantees needed, or how interested grantees would be in spending time learning from one another to identify and implement meaningful outcomes, rather than outputs, to use for program-quality improvement and reporting. As a result, the foundation expanded the coaching, community learning, and training opportunities, which have been key to the project's success.

The foundation plans to implement this project model in different issue areas, but a few key questions about the process still need to be addressed. First, how can these outcomes be sustained and what role can the foundation play to continue this progress? How can the foundation ensure that evaluative learning and program improvement via evaluation is continued, and is that possible without the main processes of CP or CPC? Second, some grantees who could have benefited from coaching, and indicated they needed individualized help, did not elect to participate at that level. How can future iterations of this work ensure understanding among the grantees about the value and purpose of coaching and overcome barriers that may hinder full participation? Third, the tension between the foundation and grantees

seeking consensus on not only reporting tools, but also the grantmaking application, was not discussed at length in this article. While the foundation will work with grantees throughout the years of a grant cycle to build capacity to improve outcomes, continued funding is always a concern for nonprofits and their programs. How do foundations balance collaboration with grantmaking deliberations? How might grantees react to such decisions, and how might it influence their continued motivation and reflection on evaluation practice?

Conclusion

The Unified Outcomes Project demonstrates the benefits that result when a foundation commits to understanding its grantees in a collaborative manner. The project helped organize outcomes within a program area that lacked unified assessment and measurement, while also encouraging grantee learning and critical reflection on evaluation practices and procedures. Although questions still exist about the sustainability of such practice, the outcomes and initial positive returns suggest that CP and evaluation coaching are powerful tools for facilitating grantee learning and reflection about evaluation practices. In addition, foundations are well situated to facilitate this type of evaluation capacity building. This process helps not only build the evaluation capacity of grantees, but may also give foundations a mechanism through which to build consensus around reporting tools to make better-informed grantmaking decisions in an open and collaborative manner.

References

- ARNOLD, M. E. (2006). Developing evaluation capacity in extension 4-H field faculty: A framework for success. *American Journal of Evaluation*, 27(2), 257-269.
- ENSMINGER, D. C., KALLEMEYN, L. M., REMPERT, T., WADE, J., & POLANIN, M. (2015). Case study of an evaluation coaching model: Exploring the role of the evaluator. *Evaluation and Program Planning*, 49, 124-136.
- GRIFFITHS, K., & CAMPBELL, M. (2009). Discovering, applying, and integrating: The process of learning in coaching. *International Journal of Evidence Based Coaching and Mentoring*, 7(2), 16-30.
- HOWARD, G. S., & DAILEY, P. R. (1979). Response-shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology*, 64(2), 144.
- KVALE, S., & BRINKMANN, S. (2009). *Interviews: Learning the craft of qualitative research interviewing*. Thousand Oaks, CA: Sage.
- LABIN, S. N., DUFFY, J. L., MEYERS, D. C., WANDERSMAN, A., & LESESNE, C. A. (2012). A research synthesis of the evaluation capacity building literature. *American Journal of Evaluation*, 33(3), 307-338.
- PATTON, M. Q. (2008). *Utilization-focused evaluation*. Thousand Oaks, CA: Sage.
- PRESKILL, H., & BOYLE, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation*, 29(4), 443-459.
- PRESKILL, H., ZUCKERMAN, B., & MATTHEWS, B. (2003). An exploratory study of the factors that influence process use. *American Journal of Evaluation*, 24(4), 423-442.
- STAKE, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage.
- STEVENSON, J. F., FLORIN, P., MILLS, D. S., & ANDRADE, M. (2002). Building evaluation capacity in human service organizations: A case study. *Evaluation and Program Planning*, 25(3), 233-243.
- TAUT, S. (2007). Studying self-evaluation capacity building in a large international development organization. *American Journal of Evaluation*, 28(1), 45-59.
- TAYLOR-RITZLER, T., SUAREZ-BALCAZAR, Y., GARCIA-IRIARTE, E., HENRY, D. B., & BALCAZAR, F. E. (2013). Understanding and measuring evaluation capacity: A model and instrument validation study. *American Journal of Evaluation*, 34(2), 190-206.
- TORRES, R. T., & PRESKILL, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, 22(3), 387-395.
- YIN, R. K. (2014). *Case study research: Design and methods*. Thousand Oaks, CA: Sage.
- Jay Wade, M.A., is a graduate student at Loyola University Chicago. Correspondence concerning this article should be addressed to Jay Wade, 111 East Pearson Street, Chicago, IL 60611 (email: jwade1@luc.edu).*
- Molly Baltman, M.A., is assistant director of grantmaking for community programs at the Robert R. McCormick Foundation.*
- Leanne Kallemeyn, Ph.D., is an associate professor at Loyola University Chicago's School of Education.*
- David Ensminger, Ph.D., is an assistant professor at Loyola University Chicago's School of Education.*
- Tania Rempert, Ph.D., is president of Planning, Implementation, and Evaluation Consulting Inc.*