Advice from local/site evaluators: How to manage "up" within a large-scale initiative

Melanie Hwalek[1], **Matt Honoré**[2], **Shavonnea Brown**[1]

[1]SPEC Associates, Southfield, Michigan, USA

[2]Oregon Health & Science University, Portland, Oregon, USA

Abstract

BUilding Infrastructure Leading to Diversity (BUILD), an initiative of the National Institutes of Health (NIH), provides grants to undergraduate institutions to implement and study innovative approaches to engaging and retaining students from diverse backgrounds in biomedical research. The NIH awarded BUILD grants to 10 higher education institutions in multiple states, including funding for local evaluations. This chapter presents findings from an online survey and interviews with 15 local evaluators from nine of the 10 BUILD sites. Participants shared their perspectives on the role of professional local evaluators in national evaluations, ideal national–local multisite evaluation partnerships, and the ways that funders can support these partnerships to maximize impact. They argued for customized technical assistance and other support for local evaluations; the importance of including local results in national evaluation findings; the value of local evaluators' subject-matter expertise; and the potential for funders to act as central organizing entities in national–local evaluation partnerships.

INTRODUCTION

BUilding Infrastructure Leading to Diversity (BUILD) is an initiative of the National Institutes of Health (NIH) that provides grants to undergraduate institutions to implement and study innovative approaches to engaging and retaining students from diverse backgrounds in biomedical research, potentially helping them to become future contributors to the NIH-funded research enterprise. NIH awarded BUILD grants to 10 higher education institutions in multiple U.S. states, including funding for local evaluations. NIH also funded the University of California, Los Angeles, to establish the Coordination and Evaluation Center (CEC), which oversees the multisite evaluation. We set out to examine the local evaluators' experiences with the CEC and other national evaluations, with a focus on what an ideal national–local evaluation partnership would look like, how professional local evaluators can enhance the effectiveness of a multisite evaluation, and what funders can do to maximize these partnerships.

Correspondence Melanie Hwalek, SPEC Associates, 313-964-0500, 18000 West Nine Mile Rd., Suite 365, Southfield, MI 48075, USA. mhwalek@specassociates.org.

RELEVANTLITERATURE

A *multisite evaluation* is defined as one that involves multiple sites and cross-site evaluation activity (Straw & Herrell, 2002). Straw and Herrell described multiple purposes for multisite evaluations: They can estimate the impact of interventions across sites, examine variation among sites, and study the impact of an intervention across a wide range of geographic locations. Multi-site evaluations are also valuable when it is necessary to acquire information rapidly, to build evaluation capacity within local communities, and/or to obtain a more adequate sample when particular population subgroups are the focus of an evaluation. Multi-site evaluations can be used to evaluate complex topics, allowing evaluators and stakeholders to work together to identify the different issues and contexts that might impact evaluation designs and findings across multiple sites. Last, funders can gain knowledge that can be used to impact policies.

Multi-site evaluations are often conducted as part of large federal or foundation grants. They typically consist of local sites that employ a local evaluator and an overarching national evaluation team. The roles and responsibilities of the local evaluator and national team can take many forms depending upon funding, size of the local team, length of the project, and requirements of the project funder. Common roles for local evaluators include data collection (Allen & Black, 2006; Chaskin, 2003; Dewa et al., 2002, 2004; Hawk et al., 2019; Niolon et al., 2016), implementation of a national evaluation plan (Biott&Cook,2000; Chaskin, 2003; Hawk et al., 2019), data monitoring, quality control, and analysis (Dewa et al., 2002; Uehara & Tom, 2011), and local evaluation capacity building (Biott & Cook, 2000; Rodi & Paget, 2007). The most common roles of national evaluators in multisite evaluations are technical assistance (Chaskin, 2003; Dewa et al., 2004; Uehara & Tom, 2011), designing a national evaluation plan and/or cross-site measures (Allen & Black, 2006; Hawk et al., 2019; Niolon et al., 2016; Rodi & Paget, 2007), data monitoring and quality checks (Chaskin, 2003; Dewa et al., 2004), and training and creation of evaluation guides for local sites (Allen & Black, 2006; Dewa et al., 2002).

Local evaluators are an asset to multisite evaluations; their unique perspective can offer knowledge and expertise that can enhance the work at the national level. Boaz and Hayden (2002) highlighted the value of local evaluators' contributions to the design of national evaluation plans and feedback on reports, which helped the national evaluators improve and refine their analyses. Allen and Black (2006) echoed these findings in their case study of the Sure Start program. Feedback sessions with local evaluators were used to confirm or check the findings of the national evaluation reports, which allowed for troubleshooting. Biott and Cook (2000) argued that the importance of local evaluators is largely dependent on the relationships they develop with the national evaluator. They drew on their own experience as local evaluators and described how the national evaluator looked to them to provide reports of local evaluation results.

The level of engagement that local evaluators have with a national evaluation depends largely on their roles. Toal et al. (2009) conducted interviews with local evaluators who had a wide range of experiences in a variety of settings and who were working on various projects funded by the National Science Foundation. They found that evaluators at the most

engaged sites participated in planning the evaluation, consulted in instrument design, and provided feedback on data collection instruments. Evaluators at less engaged sites collected data, completed annual feedback surveys, and attended trainings and meetings about data collection.

Challenges to multisite evaluations

One of the most common challenges to conducting a multisite evaluation is communication between the local and national evaluators (Dewa et al., 2004). Poor communication can lead to misunderstanding of the roles each plays with respect to the national and local evaluation. Chaskin (2003) stated that one of the challenges in multisite evaluations is the "lack of clarity regarding, goals, outcomes and expectations" at the local level (p. 75). In other words, local evaluators do not always know exactly what is expected from them. Allen and Black (2006) echoed this in their case study of the Sure Start program. There, local evaluators indicated that they did not have clear guidance on their role as it pertained to the national evaluation, and sites were not sure what to measure.

Because multisite evaluations can consist of multiple evaluators at both the national and local levels, it can be difficult to maintain consistent communication. For local evaluators, this means navigating communication not only with the national evaluator, but also clearly communicating with and meeting the needs of program staff, funders, stakeholders, and participants (Rodi & Paget, 2007). In their experience with multisite evaluation, Chaskin (2003) found it difficult to communicate equally and effectively to multiple audiences, such as funders, staff, participants, policy makers, practitioners, and academics, because each group had a different stake in the evaluation. Rodi and Paget (2007) echoed these concerns and implied that this can have ethical implications in the larger context of a multisite evaluation when the funders, staff, and participants differ in how they value the evaluation itself.

Another challenge to multisite evaluations is the increased workload that national evaluation requirements place on local evaluators (Dewa et al., 2004) as they navigate local and national evaluations simultaneously. Biott and Cook (2000) found this in their study of the National Early Years Excellence Centres Pilot Programme. High turnover of program staff, evaluation sites, and local evaluation staff can create additional challenges. Changes in program staff, particularly program leadership, can result in the loss of institutional knowledge (Rodi & Paget, 2007). Chaskin (2003), for example, found high turnover to be a barrier to multisite evaluation, particularly at the local level. (See Chapter 7 by Cobian et al. for additional discussion of this problem.) And Niolon et al. (2016) experienced the loss of a site during their 4-year multisite evaluation of Dating Matters; this had a negative impact on their ability to meet the data collection requirements of the national evaluation.

Finally, funding is a common challenge of multisite evaluations. For example, the timing of funding distribution can have major implications for completion of evaluation activities. Dewa et al. (2004) found that because funding was allocated prior to the development of a national evaluation plan, local sites had only planned and budgeted for their own site evaluations. In situations like this, funding may not be sufficient to cover the additional expenses of simultaneously participating in a national and local evaluation. Biott and Cook

(2000) also described funding challenges that arose during their participation in a multisite evaluation when a national evaluation plan was introduced on top of the local plan.

Solutions for challenges to conducting multisite evaluations

The literature offers several solutions for avoiding these challenges to multisite evaluations. Rodi and Paget (2007), for instance, suggest that if the local evaluator's role is to conduct the local evaluation and be responsible for activities set in place by the national evaluator, then the management of all evaluation activities should be a collaboration between the national and local evaluation teams. Specifically, in order to avoid miscommunication and misunderstanding of roles and responsibilities, the national evaluator and local evaluator should collaborate from the start of the project. Similarly, Chaskin (2003) suggested that national and local evaluators work together to develop "a rational and well-supported process" that establishes clear expectations for the evaluation (p. 78). This includes strategic planning, development of evaluation requirements, and identification of evaluation objectives and measures. Along these same lines, Uehara and Tom (2011) recommended that national evaluators develop an operational guide that clearly lays out the responsibilities of the local evaluator.

In Dewa et al.'s (2004) study, the national evaluator used several methods to foster better communication with local evaluators. These methods included taking advantage of email to facilitate communication among the sites and establishing a website dedicated to dissemination of and access to multisite materials, such as meeting minutes, protocol manuals, coding memos, study instruments, and proposals. The evaluator also hosted working group meetings to share major policies impacting the local sites and created a listserv and corresponding conference calls regarding data collection. Finally, the national evaluator provided physical copies of multisite materials to the local site evaluators.

To reduce the burden on local evaluators who are participating in national evaluations, Chaskin (2003) suggested clearly defining the division of labor between the national and local evaluators, establishing how instruments and data will be shared, and establishing how to conduct analyses in a collaborative manner. Dewa et al. (2004) secured additional funding for the local evaluators to assist with national evaluation activities, which allowed them to hire additional staff. Dewa et al. further noted that, to minimize the impact of high staff turnover, the national evaluator worked with the funder to provide wage increases for local staff who had taken on additional responsibilities to fulfill the requirements of the multisite evaluation.

The Role of the funder in fostering National–Local evaluation partnerships

To address funding issues, Chaskin (2003) proposed that the funder have dedicated resources to cover the expenses of both national and local evaluations. In the study by Dewa et al. (2004), once the requirements of the national evaluation were communicated, local evaluators were encouraged to review their budgets to ensure that both the national and local evaluation could be done with the existing funds. For sites that indicated the national evaluation increased their responsibilities, the national evaluator worked directly with the funder to determine additional funding needs.

Moreover, funders play a crucial role in facilitating a working relationship between national evaluators, local programs, and local evaluators. One way to build this relationship is to ensure that the national evaluation team understands what it takes to conduct a multisite evaluation. Uehara and Tom (2011) argued that funders should ensure that the national evaluation team consist of individuals with a "conceptual knowledge of evaluation [who] are competent in the basics of evaluation theories and activities" (p. 304). Funders can clearly articulate expectations regarding the roles and the responsibilities of the national and local evaluators.

Ideal models of National–Local evaluator partnerships

The evaluation literature presents some models of ideal national–local evaluation partnerships. Lawrenz and Huffman (2003) suggested that a multisite evaluation should be "objective, have the mandate of the funder, provide the opportunity for site-based stakeholders to collect and interpret data, [and] have the opportunity for sites to collaborate and develop evaluation questions" (p. 478). Their model has three stages: (a) creating the local evaluation; (b) creating the central/national evaluation team; and (c) negotiating and collaborating on the multisite evaluation.

Dewa et al. (2004) suggested a collaborative model for multisite evaluation that uses Lancaster's (1985) "six C's," which focus on designing a multisite evaluation around the following concepts: (a) contribution—each collaborator brings expertise to the project; (b) communication—having different levels of interaction as to not overwhelm or burden those involved with the evaluation; (c) compatibility—functioning as a team and appreciating the strengths of each collaborator; (d) consensus—negotiating and compromising during the course of the project; (e) credit—establishing a protocol for dissemination and authorship; and (f) commitment—providing sites with the funding and staffing necessary to meet the needs of the evaluation.

To further the knowledge base about national–local evaluation partnerships, we sought to understand what a model national–local evaluation partnership could look like from the local evaluator's perspective. Specifically, we were interested in the reflections of the local evaluators responsible for evaluating one of the 10 sites that received BUILD funding from the NIH and for collaborating with the CEC at the University of California, Los Angeles, the national evaluator for BUILD. We set out to address three questions:

1. Is there a maximally effective model for national–local evaluation partnerships?

2. What advantages and disadvantages are there to a national multisite evaluation when a professional evaluator is assigned to each site?

3. What is the role of the funder in fostering national–local evaluation partnerships?

METHODS

Each of the 10 sites funded under the BUILD initiative was expected to conduct a local evaluation as well as comply with the expectations of the CEC. These expectations were set forth in the initial request for proposals (RFP) when the BUILD sites applied for funding.

Each grantee identified a local evaluator to conduct their local evaluation and cooperate with the multisite, national evaluation conducted by the CEC.

These 10 local evaluation teams provided a convenience sample of local evaluators who all had experiences with at least one multisite evaluation. The advantage of using this sample is that all of the local evaluators were working with the same national evaluator on their BUILD evaluations. Thus, the BUILD national–local evaluation partnership served as a common context from which these local evaluators could launch broader discussions about multisite evaluations. Evaluators representing nine of the 10 BUILD sites agreed to be interviewed. One site was unable to participate because their institution stopped all involvement in research outside of the university due to the COVID-19 pandemic.

Data collection

We conducted individual Zoom interviews with one or two evaluators from each site. Each site selected which evaluation team members would be interviewed based on who had the most responsibility for conducting the local evaluation. In total, we interviewed 15 individuals. Prior to the interviews, the local evaluators were asked to complete an online survey to provide descriptive information about themselves and to offer preliminary thoughts about the relationship between their evaluations and the work of the CEC. Their responses served as a jumping off point for the interview.

The online survey asked the local evaluators to describe their evaluation team size and composition, the working relationship they had with their local BUILD grantee (e.g., external consultant, staff of the local program, evaluation unit within the university), their past experiences with other national evaluations, the length of time they had been affiliated with the local BUILD program, their role in creating their site's local evaluation plan, and the tasks they completed for the national evaluation. The survey also asked site evaluators for their opinions about what national evaluators could be providing to facilitate their local evaluation work and what local evaluators could offer to facilitate national evaluation work.

The Zoom interview questions generally followed the three study questions listed above. We collected perceptions and opinions about how national and local evaluators can work together to facilitate data collection, the possible roles of funders, how local evaluators can contribute to a multisite evaluation, how a national evaluator can support local evaluators, and what an ideal national–local evaluation partnership would look like. The interviews also inquired about the local evaluators' experiences with other national evaluations where they served as either a local evaluator or a national evaluator. The intent of this broadened inquiry was to obtain as many thoughts as possible about national–local partnerships and not solely constrain comments to participants' experience with BUILD and the CEC.

Two interviewers from the study team conducted each interview. The interviewers were, themselves, local evaluators for one BUILD site, which gave them a deep understanding of the roles and responsibilities of the local evaluators regarding the expectations of the BUILD national evaluation. The study team also completed the online survey and answered the interview questions themselves.

Data analysis

We aggregated the online survey responses for the nine institutions using Microsoft Excel. For two institutions, two evaluators completed the online survey separately. Since the online survey was about the evaluation team as a whole, we used the primary evaluator's answers to have one set of survey responses for each of the nine institutions. We generated descriptive statistics from the responses to the online survey.

We analyzed the interview responses in two ways. Sobo et al.'s (2003) method for rapid assessment of qualitative data from telephone interviews was our first step. In this method, analysis happens as the interviews take place. Interviewers take detailed notes to identify patterns and hypotheses, generate direct quotes, and paraphrase material as needed. After each interview, the interviewer summarizes their notes, then the interviewers come together to review their notes and agree on themes. If there are differences in findings, the interviewers discuss them until there is agreement. This use of rapid assessment methodology allowed the interviewers to explore any hypotheses or themes that emerged from prior interviews during subsequent interviews. One adaptation we made to this method is that we did not capture quotes during the debrief sessions because the interviews were recorded and transcriptions of exact quotes could be generated later. We created transcriptions of the recorded interviews using either REV.com or GoTranscript.com.

The second method of data analysis was coding the interview transcripts using QDA Miner software. The initial coding scheme was structured to align with the study questions and themes that came from the rapid assessment discussions. As coding proceeded, we added new codes and subcodes as needed. One interviewer coded all transcripts and then abstracted the quotations, organized by code, into a table. The second interviewer reviewed the table of abstracted quotations and validated that the quotes did, in their assessment, reflect the codes and subcodes to which they were assigned.

Study respondents

The local evaluators who were interviewed varied in their prior experiences with national evaluations and in the length of time they had been involved with their BUILD sites. At six of the sites, at least one evaluator on the team had been the evaluator since the site received its BUILD grant. At five sites, the evaluators wrote the original local evaluation plan. Evaluators at five of the sites had prior experience as national evaluators. Evaluators at six of the sites had experience as site evaluators in other national evaluations. The number of full-time equivalents (FTEs) that the local evaluators assigned to their BUILD evaluations ranged from one to four. Six of the sites had three or four FTEs currently working on their local evaluations.

At three sites, the local evaluator was on the staff of an evaluation unit within their university; three were external evaluation consultants. Three were staff or faculty at the institution that received the BUILD grant. Regarding their responsibilities in the national evaluation, four of the sites were responsible for uploading participation data to the CEC's Tracker database; six were liaisons with the CEC for the national longitudinal surveys; and five provided institutional research data. When local evaluators' roles were disaggregated by

their relationships to their site, it was discovered that none of the external evaluators reported being responsible for uploading participation data into the Tracker database or for providing institutional research data to the CEC. Only one external consultant reported being a liaison to the CEC for the national surveys. Table 1 shows these results.

RESULTS

Is there a maximally effective model for National–Local evaluation partnerships?

Participants were asked two main questions: What does a maximally effective national–local evaluation look like? And what helps or hinders partnerships? Their answers to these questions heavily overlapped. Three basic themes emerged regarding what could support an effective partnership:

- a national evaluation team configured to also support the local evaluations;

- utilization of local evaluation findings in the national evaluation; and

- technical assistance from the national team in instrument design and data analysis.

A National evaluation team configured to support the local evaluation—

Respondents suggested that it would be useful if the national evaluator had someone from the national evaluation team embedded directly on the local evaluation team. The national evaluation team member could directly offer their expertise and, in turn, would have a better understanding of the local context:

> The ideal relationship would be the [national evaluation team] would be a coordinating and assistance center that would help in having a representative that is more directly engaged with the specifics of your program…in the sense that they're immersed in it.…That might be better housed at a national level, but [they] really are there at the service of the local sites and are there to support the local sites in developing strong programs and promoting strong programs and effective evaluation.

Respondents shared that the benefits of immersion could be to optimize collaboration, provide a level of support that reduces the burden for local evaluators, and identify the unique aspects of each site. For example:

> They would essentially become immersed in one, two, or three sites. And they would just really have to just become immersed in what that culture is and what the environment is for that site. Just become an expert in one or two or three sites and be integral to the team. Be the representative from the coordinating center to advocate and serve the local sites.

> I'm thinking about the arguments that people have made for multicultural validity. I wonder if you could go along the same lines here in terms of saying, "If you understand the sites better, you will be able to report more accurately on findings."

Another way respondents expressed that the national evaluator could support the local evaluators is to pay attention to making the additional work of collaborating on the national evaluation feel less burdensome:

> [A national evaluation team] should have such good communication that it knows when it's making it harder and it knows when it's being helpful.…The relationship would be that the coordinating center is in service to the work that's being done at the sites and helps to create collateral benefit from the work that's happening at the sites and not creating more work.

Respondents commented that the configuration of a national evaluation team can be beneficial or detrimental to meeting the needs of local evaluators. Too many individuals in leadership positions and not enough people on the ground can make timely service to local evaluation teams difficult. For example, one local evaluator described the national evaluation team as "way too top-heavy." More precisely, this evaluator felt that the team was "really under-resourced in terms of research assistants and project managers and people to actually do on-the-ground work."

Utilization of local evaluation findings in the national evaluation—Respondents suggested that national–local evaluation partnerships would maximize their effectiveness if site-level learnings were fully integrated into the national evaluation findings. They believed that this could make the national evaluation more meaningful.

> It might be nice to see [a model] where the national evaluator doesn't necessarily come in with a [longitudinal survey] that's already developed and conceptualized, but instead takes data from all the different sites and says, "What can we do with this? What are we learning from the ground up? And how can we really use the data that's coming from campuses to tell a story?"

Respondents noted that typical meetings with the national evaluator were process-oriented, focusing largely on hearing about how things were going, particularly with respect to data collection for the national evaluation. There was a desire for more meaningful engagement about what was being learned from the evaluation. For instance, one participant called phone calls with the national evaluator "just a one-sided push-out of information, and not a dialogue." They thought these exchanges would be more productive if they had been "working meetings and some of them had been dialogues, rather than just reporting out.… Those could be an email." Another participant described a similar scenario:

> In large part, it is the site evaluators providing the national evaluators with information that they require and need for their purposes with more limited interest and attention focused on what is the value [of the local evaluation] to learning at the site.

Provision of expertise in instrument design and data analysis—Respondents mentioned that the expertise of the national evaluator is an important component of an effective national–local partnership. In the case of BUILD, local evaluators specifically mentioned that the CEC's psychometric expertise could add value to the local evaluation in terms of instrument development. For example:

> [The CEC has] a strong skillset in instrument development.…The CEC is really good at developing items that actually measure what you're trying to measure. They have that horsepower, and I've always appreciated that.

> These are social behavioral scientists who are very, very knowledgeable in that particular area—science identity and self-efficacy and stuff like that. I'm sure that they are better versed in all that than I am.

The respondents also mentioned other types of technical assistance, including showing local evaluators how to make use of national evaluation data and offering support for more complex data analysis. One interviewee provided a specific example:

> What would be useful would be a situation whereby, let's say, I know that I need to do propensity score matching. I've never done propensity score matching. I contact someone [from the national evaluation team] and say, "We need some help with propensity score matching." Then there's an actual person who can sit down and work through those needs with us.

Respondents mentioned the value of having a national data collection platform that the local evaluators could use to provide clarity on the meaning of measures, offer a standardized methodology for constructing measures from items in a national data set, and store all instruments. The CEC has created such a model for BUILD, but respondents took this a step further:

> [The] CEC needs to play a very key role here in either monitoring, supervising, or doing that part of the project—the data analysis. Database building, like, "This is my research question. What should I do?" Then the CEC can say whether they have the data or not, what data needs to be in there, what confounders need to be in there, rather than each group deciding on their own.

What advantages and disadvantages are there to a national multisite evaluation when a professional evaluator is assigned to each site?

Respondents did not envision any disadvantages to a national evaluation with professional evaluators at the local level. They identified two key advantages to having these professional evaluators: their own subject-matter expertise and their ability to bring the local perspective to the national evaluation.

In the case of BUILD, evaluators noted many of the local site principal investigators had backgrounds in the hard sciences (e.g., chemistry, biology) and the BUILD intervention itself was grounded in social science concepts. As such, professional evaluators—many of whom have backgrounds in the social sciences—are suited to oversee evaluation at the local level:

> Should social scientists be at the helm of a social science project, or should physical scientists be at the helm of that? I think for this kind of question, you probably want to engage the experts who have been thinking about this question for a long time.

> I think the danger is that the [principal investigators] have—they're mostly people who have experience running student training programs. Most of them are not researchers. They don't do educational research. They're biologists and chemists and people who run training programs, and the evaluation is on the side.

While a national team can be invaluable in the creation of a design for use across multiple sites, local evaluators give a perspective on how that design will look in their local contexts. In the case of BUILD, consortium-wide Hallmarks of Success were created nationally with input from local stakeholders. These Hallmarks acted as outcome indicators from which the CEC created measures. The CEC sought input about the Hallmarks and their measurement, but some local evaluators reported it was their BUILD principal investigators who received the invitation for feedback; the invitation did not necessarily filter down to the local evaluators. Respondents also felt that more attention could have been paid to what could be validly measured in local contexts:

> [Principal investigators] definitely had more of a say in these [Hallmarks]. And then it was fed down to evaluators, but I don't think it was that communicative. It was definitely more about, "These are the Hallmarks," as opposed to what kind of data can you collect easily or difficult.

What is the role of the funder in fostering National–Local evaluation partnerships?

Respondents drew on their experiences with BUILD and other national evaluations to provide recommendations for how funders could foster effective relationships between local and national evaluators. The most common suggestions were:

> articulating the roles, responsibilities, and expectations of both the national and local evaluators before the multisite evaluation is launched; acting as the central organizing entity, keeping both the national and local evaluators focused on the purpose of the evaluation; and playing a major role in easing the tension of grantees not wanting to share negative results.

Articulate expectations early—Funders can make clear the exact nature of what the working relationship will be between national and local evaluators. One respondent said that they have "never seen a national evaluation or RFP that really defined enough how much the sites will need to be involved in the national evaluation." They explained that "that's always been a disconnect. I have not seen that done well yet and that's problematic." Indeed, respondents noted their desire to have these roles clearly defined early and as part of the RFP process. For instance, one interviewee described the frustration that came out of an initial lack of clarity:

> We all thought we were doing our own evaluations.…Where the funder comes in, it was implicit that our funding might depend on us cooperating with the national evaluation. We always felt that we had to do the logic model, to do everything they wanted, which took tens of hours of work that was not planned and not funded so that the national evaluation could have some legitimacy.…I think [the national evaluator] got funded really late.

Sequencing the timing of evaluation planning between the national and local evaluator is another way that funders can enhance working relationships in a multisite evaluation. Respondents mentioned the need for adequate planning time prior to the launch of the evaluation to allow for instrument tuning and local institutional review board (IRB) submissions:

> I think one thing is to allow adequate timing for all of the steps that need to happen. For example, we needed to submit through our local IRB.…[The national evaluation team] would come up with their instrument designs, pass it around for review, and sometimes send it to their IRB and get it approved and then still be making changes to it while we needed to send it to our IRB.…It's closer to, I think, a month process for [our local IRB] to review and approve a submission. Things get backed up.

Another participant said it would be helpful to have the national evaluation team set up first so that they could then "write the criteria for the evaluation design…so that all the sites know that the [national evaluator] is centrally involved from the beginning."

Absent a clear understanding of what kinds of information would be collected and shared and by whom, the funder, national evaluator, and local evaluator might create overlapping timelines and instruments designed to collect the same information. These funder reporting requirements and national evaluation survey priorities often create undue burden for local data collection and can cause survey fatigue among students:

> Survey fatigue was a big one. [We had to learn whose data collection was most important].…A lot of times we were stuck, couldn't move without sacrificing something like our own spring data collection or saying, "Wow, they're not going to change their timeline. And [the national evaluator] is aware of [the funder's] timeline [and] that they're all asking the same thing." But no one's budging from how they gather their data.

> Our local survey, they're filling out science identity. They're filling out research self-efficacy questions. They're filling out leadership questions. They're saying, "How often do they do research?" and then they do it all again, on the national survey.…I do think that what would be helpful is that local sites could do that. That we could funnel, particularly if we're doing our first funnel I gave to them, and then there would be a smaller burden on the students in terms of the second survey, and a smaller burden on the national group for collecting data.

According to respondents, the breadth of responsibilities for local evaluators is not always clear, particularly when it comes to knowing what information would be useful for the national evaluation and how to handle data analysis. As one participant explained:

> Do you expect the local sites to be producing high-level statistical analysis? Then you need to say that from the beginning, so that they can have the right people on board and collect the right data in the right way.

Respondents commented that funders can articulate in RFPs what the interactions between national and local evaluators will be, and toward what purpose. They can be explicit about

the expected relationship by requiring that funding be allocated specifically to national–local collaborations:

> Suppose they had a model that said, "We're giving you X number of dollars to spend on a local evaluation." [You] would set a percentage. "We want you to set that aside for your local evaluation, and we want the local evaluators to commit to coming together to figure out their plan in the first year."

The complexity of balancing this relationship is explored further in Chapter 1 by Guerrero et al.

Act as a central organizing entity—Funders have the ability to keep a focus on the overall purpose of an evaluation. They can, for example, establish an overarching evaluation question, clarify what local evaluations should measure or are better at measuring, articulate the expected working relationship between national and local evaluators, and sequence the timing of evaluation planning. One respondent tied this issue to their own teaching:

> These are the things that we're trying to teach to students—how to develop a research question; you need to look at the gap in the literature; contribute to filling the hole of our knowledge. We're not doing that. We don't even know what the gap is.

Another said that it would have been useful for local sites to see the proposal that funded the national evaluation. This would have provided insight into "maybe why it got funded." They explained that it would have been useful for the funder to explain, "This is why we funded this proposal. This is where we see it working, and this is what we're expecting."

In fact, respondents suggested that funders talk directly with local evaluators prior to establishing a national evaluation team in order to understand what should be measured at the local level. Local evaluators bring rich understanding of their sites and may have measures that they already use in their own populations. Doing so would allow site evaluators to share "their expertise and [it means] you're getting their knowledge of their sites, that they're bringing that with them into the design of the national evaluation measures, and maybe even methods."

Ease the angst of publicly sharing negative results—Funders can play a key role when it comes to sharing what doesn't work. Respondents commented that there was perceived pressure from their grantees not to publicize failures. One explained that "you don't want to share your data if it's bad, because you're afraid you're going to lose your funding in the second year [before you can] get the kinks worked out. Well, it's never enough time to get the kinks worked out." Another respondent made a similar point:

> That tension between proving things work and wanting to learn. It probably comes all the way down from [the lead funding organization]… because [the funder] is like, "What can we tell them to make them fund us for another 5 years?" That's all about success to me.

DISCUSSION

The working relationship between a national evaluator and local evaluator can have major implications for a multisite evaluation. The literature on multisite evaluations highlights the challenges and solutions to national–local evaluation partnerships, the importance of local evaluators, ideal national–local evaluation partnership structures, and the role that funders play in fostering national–local evaluation partnerships. This chapter looked at these issues as expressed by BUILD local evaluators, many of whom had prior experiences working as local and/or national evaluators on other multisite evaluations.

Local evaluators see the burden of national evaluation requirements falling not only on local evaluation teams but also on study participants; students in many BUILD sites had to answer duplicate questions on the national and local surveys. Future research should explore whether this survey burden has a ripple effect on students' willingness to participate in other aspects of the evaluations, such as focus groups, site visits, or oral histories.

This study further uncovered that a key contribution of local evaluators is their ability to help national evaluators understand the variations that are likely to occur across sites within a national database. Kirkhart (2010) spoke of the importance of multicultural validity and the need to examine theory within the cultural context. While national survey data can be disaggregated and results can be compared across sites, they cannot capture culturally specific reasons that may be behind any variations. Incorporation of data being collected by local evaluators, when designed properly at the beginning of a national evaluation, can offer rich insight into a national evaluation, especially when there is wide variation across sites.

The BUILD initiative is quite different from many of the studies cited in the literature, most of which reported that all sites were implementing the same or very similar interventions (Dewa et al., 2004; Niolon et al., 2016; Rodi & Paget, 2007; Uehara & Tom, 2011). Given the diversity in the types of interventions being implemented and in the institutional contexts among the 10 BUILD sites, the BUILD initiative is, as one respondent noted, essentially 10 different experiments:

> What's beautiful about BUILD is you had 10 experiments. I think that's a really amazing design. It's like around the world in 80 days of research. Who's going to crack this nut first based on our different approaches? We also are dealing with unique spaces and grounding that are representative and could have lots of implications for other places.

The importance of multicultural validity does not preclude the importance of using common local evaluation measures across the sites. Richer data for a national comparison could have been gathered had the local evaluators, for example, decided together on a common set of interview or focus group questions to ask students. Meetings among local evaluators across the sites after a study begins cannot be expected on their own to generate common measures across sites. This requires facilitated discussions between national and local evaluators from the beginning of the initiative, with a goal of ensuring multicultural validity of the national evaluation findings. Study respondents alluded to this possibility in their desire to make meetings with national evaluators more mutually beneficial and learning oriented.

This study identified many factors that facilitated an effective multisite partnership, including receiving technical assistance and having the national evaluation also be of service to local evaluators. Previous literature on ideal national–local evaluation partnerships suggests that for partnerships to be effective there must be collaboration between national and local evaluators at every stage, from planning to dissemination (Dewa et al., 2004; Lancaster, 1985; Lawrenz and Huffman, 2003). Study respondents went so far as to suggest that the national evaluator could be integrated directly into local evaluation teams. In this way, local evaluators could receive just-in-time and customized advice as their evaluations unfold while simultaneously giving the national evaluator a rich picture of the local context.

This study, coupled with the evaluation literature, suggests that effective models for national–local evaluation partnerships can vary and that they seem to depend on the degree of similarity across sites. A top-down approach, where the national evaluator is fully in control and offers little to no opportunity for collaboration with local evaluators, may work better when interventions are standardized and common assessment tools are in use. A bottom-up approach, in which the national evaluation is completely informed by local evaluation data, could also work if there are professional evaluators capable of validly collecting data beyond monitoring, and if there is sufficient time for local and national evaluators to collaboratively design the national evaluation plan and agree to common measures.

There is also a middle model that offers true collaboration, where both the national and local evaluators are contributing their own subject-matter expertise, suggesting standardized measures, and collectively interpreting findings. Our study suggests that this middle model would be most effective for BUILD. The CEC and the NIH provided Hallmarks of Success that guided the conceptualization of outcomes across the 10 sites while enabling each site to define and deliver its uniquely designed interventions—in essence, "10 beautiful experiments."

Had this middle model been adopted in the planning phase of the national BUILD evaluation, it could have resulted in a very different evaluation design. For instance, had local evaluators been given the opportunity to collaborate on the national evaluation design, their reported subject-matter expertise suggests that the national evaluation may have placed more emphasis on examining deficit and non-deficit models of intervention across sites and on gathering more qualitative information about students, institutions, and the communities within which they reside. This is confirmed in the literature: Boaz and Hayden (2002) and Allen and Black (2006) both discussed the important contributions local evaluators make to multisite evaluations by using their expertise to inform national evaluation plans, design instruments, and provide feedback to the national evaluator.

The relationship that local evaluators had with the CEC did not seem to differ based on whether the site evaluator was an external contractor, staff of an evaluation unit within the university, or a faculty or staff member at the institution. All evaluators indicated similar experiences in their working relationship with the CEC. The literature indicates that the role of the local evaluator is dependent upon the needs of the sites (Toal et al., 2009). There was

no mention in the literature of the working relationship being impacted by the particular expertise of the local evaluator or their structural relationship with the site.

Funders may be able to foster the relationship between national and local evaluators by acting as a centralized organizing entity and by articulating expectations early. Previous literature has called on funders to clearly articulate their expectations for the roles and responsibilities of national and local evaluators (Rodi & Paget, 2007). One way to communicate the importance of local evaluations is to set guidelines for their budget in the RFP. This can help grantees understand their importance to the success of the intervention as well as ensure that local evaluators have the resources they need to conduct high-quality local evaluations and support the national evaluation. Past literature on multisite evaluations does not address the potential role funders can play by placing value on the local evaluation or by providing budget guidance.

Last, funders can play a role in easing the tension of grantees who do not want to publicly share negative findings—a concern expressed by several study respondents. Funders can help reduce this angst by explicitly stating in the RFP that sharing negative findings is a requirement and by hosting national convenings explicitly to share failures. There is a gap in the multisite evaluation literature regarding what role funders can play in bringing negative findings to the forefront. This does a disservice not only to the national evaluation but also to society at large. As Martin Luther King (1968) said in his Distinguished Address to the American Psychological Association, it is a role of the social scientist to "tell it like it is."

AUTHOR BIOGRAPHIES

Melanie Hwalek, PhD, is chief executive officer of SPEC Associates and assistant professor of psychology at Michigan State University.

Matt Honoré, MPA, is a research data analyst at Oregon Health and Science University, where he has been a part of the BUILD-EXITO evaluation team for the past 6 years.

Shavonnea Brown, MA, is an evaluation specialist at SPEC Associates, a nonprofit that specializes in creating tailor-made program evaluations that emphasize learning.

REFERENCES

Allen M, & Black M. (2006). Dual level evaluation and complex community initiatives: The local evaluation of Sure Start. Evaluation, 12(2), 237–249. 10.1177/2F1356389006066974

Biott C, & Cook T. (2000). Local evaluation in a national Early Years Excellence Centres pilot programme: Integrating performance management and participatory evaluation. Evaluation, 6(4), 399–413. 10.1177/2F13563890022209398

Boaz A, & Hayden C. (2002). Pro-active evaluators: Enabling research to be useful, usable and used. Evaluation, 8(4), 440–453. 10.1177/2F13563890260620630

Chaskin RJ (2003). The challenge of two-tiered evaluation in community initiatives. Journal of Community Practice, 11(1), 61–83. 10.1300/J125v11n01_04

Dewa CS, Butterill D, Durbin J, & Goering P. (2004). No matter how you land: Challenges of a longitudinal multi-site evaluation. The Canadian Journal of Program Evaluation, 19(3), 1–28.

Dewa CS,Durbin J,Wasylenki D,Ochocka J,Eastabrook S,Boydell KM,& Goering P. (2002).Consideringa multisitestudy? How to take the leap and have a soft landing.JournalofCommunityPsychology,30(2), 173–187. 10.1002/jcop.10001

Hawk M, Riordan M, Fonseca JJ, & Maulsby C. (2019). I don't want the tray to tip: Experiences of peer evaluators in a multisite HIV retention in care study. AIDS Education and Prevention, 31(2), 179–192. 10.1521/aeap.2019.31.2.179 [PubMed: 30917016]

King ML Jr. (1968). The role of the behavioral scientist in the civil rights movement. The American Psychologist, 23(3), 180–186. 10.1037/h0025715 [PubMed: 5643229]

Kirkhart KE (2010). Eyes on the prize: Multicultural validity and evaluation theory. American Journal of Evaluation, 31(3), 400–413. 10.1177/2F1098214010373645

Lancaster J. (1985). The perils and joys of collaborative research. Nursing Outlook, 33(5), 231–232. [PubMed: 3849721]

Lawrenz F, & Huffman D. (2003). How can multi-site evaluations be participatory? American Journal of Evaluation, 24(4), 471–482. 10.1016/j.ameval.2003.09.003

Niolon PH, Taylor BG, Latzman NE, Vivolo-Kantor AM, Valle LA, & Tharp AT (2016). Lessons learned in evaluating a multisite, comprehensive teen dating violence prevention strategy: Design and challenges of the evaluation of dating matters: Strategies to promote healthy teen relationships. Psychology of Violence, 6(3), 452–458. 10.1037/vio0000043 [PubMed: 29607239]

Rodi MS, & Paget KD (2007). Where local and national evaluators meet: Unintended threats to ethical evaluation practice. Evaluation and Program Planning, 30(4), 416–421. 10.1016/j.evalprogplan.2007.06.005 [PubMed: 17945142]

Sobo EJ,Simmes DR,Landsverk JA,&Kurtin PS(2003).Rapidassessmentwithqualitativetelephoneinterviews: Lessons from an evaluation of California's Healthy Families program & Medi-Cal for children. American Journal of Evaluation, 24(3), 399–408. 10.1177/109821400302400308

Straw RB, & Herrell JM (2002). A framework for understanding and improving multisite evaluations. New Directions for Evaluation, 94, 5–16. 10.1002/ev.47

Toal SA, King JA, Johnson K, & Lawrenz F. (2009). The unique character of involvement in multi-site evaluation settings. Evaluation and Program Planning, 32(2), 91–98. 10.1016/j.evalprogplan.2008.10.001 [PubMed: 19062092]

Uehara DL, & Tom T. (2011). The context of evaluation: Balancing rigor and relevance. Journal of MultiDisciplinary Evaluation, 7(15), 296–305.

TABLE 1

Local evaluator responsibilities in the national evaluation by relationship to their site

Tasks	Evaluation unit within a university ($n = 3$)	External consultant ($n = 3$)	Staff of faculty of site grantee ($n = 3$)	Total ($N = 9$)
Tracker roster uploads	2 (66%)	0 (0%)	2 (66%)	4 (44%)
Liaison for national surveys	3 (100%)	1 (33%)	2 (66%)	6 (66%)
Provide institutional research data	2 (66%)	0 (0%)	3 (100%)	5 (56%)