

The Context of Evaluation: Balancing Rigor and Relevance

Denise Lea Uehara

University of Hawaii Manoa-CDS

Tammy Tom

University of Hawaii Manoa-CDS

Background: The context in which an evaluation is undertaken impacts not only the core evaluation activities but also the ethical standards that guide our work. Theoretical constructs and ethical decision-making frameworks often may not support us in ethical dilemmas given certain evaluation settings.

Purpose: The purpose of this article is to present an example of how a well-intentioned, responsive yet rigorous evaluation provided opportunities to expand our experiences. We conclude this paper with a set of recommendations for organizations/institutions, evaluators, project developers/ implementers, and grantors

Setting: Not applicable

Intervention: Not applicable.

Evaluating complex interventions is complicated. Evaluating an under-developed complex intervention within a public/private partnership (PPP) while maintaining scrupulous evaluation practices is even more difficult. Couple this situation with both the independent evaluation and the project housed in the same organization. Efforts to execute a successful evaluation amidst a sometimes volatile atmosphere provide valuable learnings about being responsive yet

Research Design: Not applicable.

Data Collection and Analysis: Desk review.

Findings: The complexities of conducting an external evaluation of a study that relied on a public-private partnership coupled with a complex intervention that was not fully developed resulted in unintended study limitations and necessitated adapting evaluation strategies and making ad hoc adjustments midstream. Through our recommendations we share our lessons learned and how challenges can be creatively and ethically addressed at the onset of an evaluation.

Keywords: *multi-tiered evaluation; evaluation context, ethical standards, evaluator role*

rigorous. The context in which an evaluation is undertaken impacts not only the core evaluation activities but also the ethical standards that guide our work. We offer our responses to these challenges as recommendations for various role groups and define evaluator flexibility, as we perceived it to be in this setting.

This paper extends the discussion first highlighted by Chaskin (2003) then expanded by Rodi and Paget (2007) about unanticipated ethical threats associated

with multi-tiered evaluations. We contribute to the conversation by adding a complex intervention, public/private partnership and conflict of interest to the multi-tiered evaluation context.

We confirm the opinion of Rodi and Paget when they refer to the Joint Committee on Standards for Educational Evaluation (Fitzpatrick, Sanders, & Worthen, 2004) and the *Guiding Principles for Evaluators* (American Evaluation Association, 2004). Sometimes, theories and principles governing our work may be inadequate guides. Their article illustrates situations where theoretical constructs and ethical decision-making frameworks often may not support us in ethical dilemmas given certain evaluation conditions, particularly conditions involving multi-tiered projects with multi-tiered evaluation structures. They define multi-tiered projects as those that are implemented in many political settings with separate managerial and organizational arrangements operating under the same program rubric and responsible for a collective set of goals and objectives. Multi-tiered evaluation structures include layers of evaluation at the state or local and national level. These multi-tiered approaches create a number of problems for ethically executing evaluation activities. Rodi and Paget suggest applying an appraisal of power relations to address conflicts inherent in complex evaluation circumstances. Their evaluation ethic was to reduce power differences among evaluation stakeholders while ensuring that activities adhere to professional standards. In our particular situation, power differentials fluctuated across partners, grantees, evaluators (local and national) and grantors. Based on our context (under-developed complex intervention within a PPP, project and evaluation teams

employed in the same organization) briefly described earlier, we were forced to define our roles not in terms of reducing power, but establishing, diffusing, and deflecting it. In other words, the roles and functions of an independent evaluation were constantly tested by all parties.

Conceptual Framework

The context is situated from the perspective of an independent evaluation working within the same organization as the project developers/implementers. "Independence" was defined as having mutually exclusive project and evaluation functions and activities, including separate budgets and supervision of staff. Our evaluation reflected two purposes not uncommon to most evaluations: (1) conduct a rigorous evaluation—provide empirical answers for legislation regarding health care (2) employ a relevant framework—provide useful information for consumers. We present an example of how we responded to unexpected challenges arising within a well-intentioned, responsive yet rigorous evaluation.

Project Description and Evaluation Context

The case presented here is a federally funded twelve-month community-based study testing an intervention intended to delay or prevent diabetes-related disability, unemployment, and reliance on federal assistance programs. The study contract was awarded to a state agency in 2006 and required a rigorous independent evaluation utilizing a randomized controlled trial (RCT) design. The study involved a partnership between 1) a large state government agency

comprising two separate entities (a university and a state department) and 2) a small privately incorporated business targeting health and wellness among local employer groups. The authors of this paper serve as members of the independent evaluation, funded by the federal agency under sub-contract with the state grantee. The federal agency had also contracted with a large private company to serve as the multi-site evaluator contracted to provide technical assistance to and develop reports for the federal agency.

Both the project and evaluation staff was employed by the same organization. This conflict of interest circumstance was described by Gorman and Conde (2007) in which secondary competing interests influenced professional judgment related to research validity. Although specific to the school-based drug and violence prevention literature, their investigation into the scope of the problems pertaining to competing interests between program developers and evaluators, reinforced the need for evaluation separation and independence, particularly when assessing intervention effectiveness.

Project Structure and Partnership Description

In 2006, a state agency received federal funds to determine/demonstrate the value of providing health care benefits and other services to support individuals at risk of becoming disabled. A cohort of four states was funded contingent upon

implementing an independent evaluation utilizing experimental design that would collect uniform data for cross-site comparisons for a private research company. The federal funding agency contracted with this private research company to conduct the federal-level evaluation. With research unfamiliar to the state agency who traditionally received funds for service delivery, the state agency sub-contracted the planning, implementation, and required independent evaluation of the project to a state university unit. The state agency had little if nothing to do with the project and evaluation activities.

In developing the grant proposal, the university discovered that a private business organization was providing services to the target population. This business offered its member companies and their employees services in health and drug plan benefits, health plan disease management assessment, education related to the healthcare industry, and employee worksite wellness. Discussions were held with the organization and its members and several member companies expressed interest in participating in the study. Given this information, the project decided to enter into a partnership with the private organization to conduct the study. Figure 1 illustrates the overall structure and general responsibilities of the funder, contracted multi-site evaluator, state grantee, and partner organization.

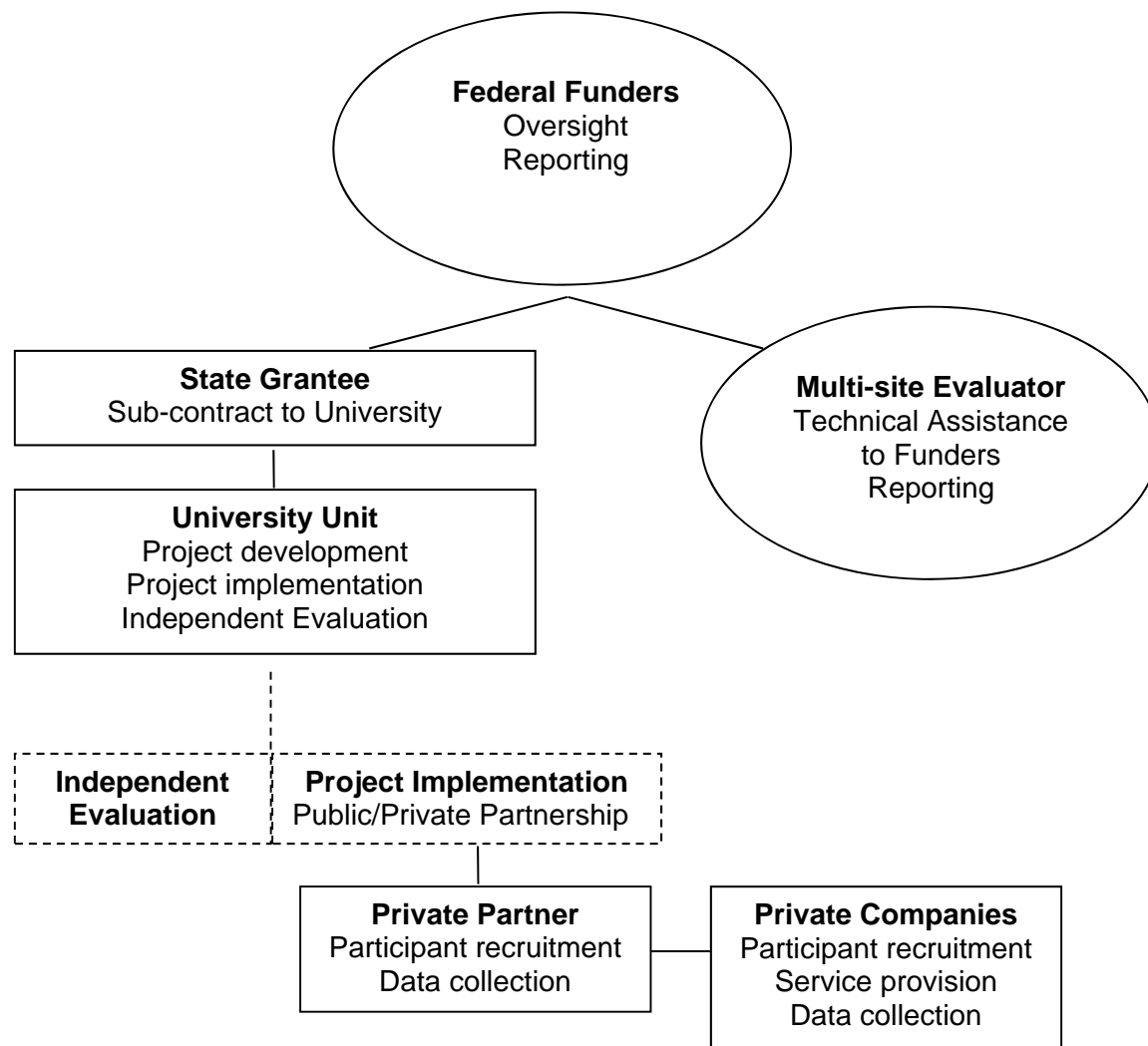


Figure 1. Model of Study Entity Roles and Responsibilities

Intervention Complexities

One of the difficulties surrounding the evaluation of this project was the complexity of the intervention itself and how the various parts interacted within the PPP. Complex interventions typically contain several interconnected components consisting of numerous characteristics. Complex interventions can be characterized by:

Number of interacting components within the experimental and control interventions; Number and difficulty of behaviors required by those receiving the intervention; Number of groups or organizational levels targeted by the intervention; Variability of outcomes; Degree of flexibility or tailoring of the intervention permitted (Medical Research Council, 2008, p. 7).

A successful evaluation of a complex intervention requires understanding how to deal with these complexities. Therefore, it is critical to understand the range of potential intervention effects and how they may vary and inter-relate. It is essential for the evaluator to understand how the intervention works, what the key ingredients are and how they demonstrate effects. Much of this knowledge is traditionally understood by the parties implementing the study. However, for this particular study, the intervention was often modified during implementation of the RCT design. Furthermore, because of the developmental nature of the intervention, the RCT design was not compatible yet required.

Contextual Challenges

The partners had little if any research experience implementing an experimental design with random assignment. They were committed to providing services rather than committed to the funded and required evaluation. Standardization, model definition and model adherence documentation was important in order to provide an accurate description of the intervention. The evaluation made several recommendations pertaining to these concerns.

The partnership itself had many challenges further complicating the evaluation activities. While the funders applauded the PPP for being innovative and progressive, the lack of follow through and inconsistent quality of executing their respective partnership responsibilities often threatened the evaluation. Evaluation rigor was frequently perceived as unimportant or inconsequential by project staff and especially the private partner who was

also responsible for critical data collection activities.

Considerably evident throughout this unstable climate was the absence of clear lines of authority and communication from the funders and the multi-site evaluators. For example, the multi-site evaluators always prefaced their suggestions with, "We are agents of X and cannot speak for them." Yet when directly seeking guidance from the federal program officer, we were consistently directed back to the multi-site evaluators. Furthermore, the project had gone through three federal program officers within two years. Needless to say lines of accountability and adequate support for project and evaluation implementation were sorely wanting. We share the aggravation reported by both Chaskin (2003) and Rodi and Paget (2007) about the "inconsistent and diverse identities of the evaluation client and consumer." For example, threats to the RCT were brought to the attention of the funders prior to and periodically throughout the participant recruitment period-knowing that recruitment efforts would not yield the targeted sample size. Requests were made to re-design the evaluation (quasi-experimental, developmental, descriptive, success case method) to accommodate the potential threats. However, the funders, more specifically their contracted evaluators, strongly encouraged us to retain the RCT design.

There were several reasons why we were asked to continue with the RCT: 1) it was required and approved for funding (despite not being appropriate for the project) 2) the other three states who received funding prior to us already implemented their RCT designs 3) the data collected for this design supported the universal data set of the multi-site contracted evaluators. The RCT proceeded

with 35% of the targeted sample enrolled and the blessings of the funders. Given this situation, we positioned the evaluation to offer meaningful findings by distributing power amongst all players.

Influence of Contextual Challenges on Evaluation

A major concern identified at the project's inception was having the project independently evaluated by staff housed within the same organization. The organization typically conducted internal evaluations of their own projects, with a project director supervising both project and evaluation staff. In this situation, although the independent evaluation functioned apart from the project and included a separate budget and supervision, the project director often perceived the role as supervising both project and evaluation staff which resulted in conflict and compromised relationships. In some instances, fidelity monitoring data, collected in a confidential manner and intended only for the purpose of the independent evaluation, was requested by the project for use in intervention staff personnel evaluations. Additionally, the evaluation's efforts to disclose truths or flaws in the study conflicted with the project teams' advocacy of the project. Formative evaluation activities designed to provide feedback for improvement purposes were often met with a defensive stance. There was an unspoken expectation from both public and private partners to produce results that would make the intervention appear successful rather than to produce results from a successfully executed evaluation. These subtleties made communication between the project and evaluation difficult.

Private partner actions at times appeared to be in direct conflict with project goals. For example, recruitment by the private partner into similar programs was conducted simultaneously and with the same participant pool that was planned for the project. This limited the participant recruitment pool within the private partner companies. These types of conflicts were not addressed in MOAs and therefore could not be referred to as a breach of contract if discussed.

One of the major responsibilities of the private partner was to provide intervention services, which included documenting interactions with project participants. This type of data related to the measurement of fidelity and informed a comprehensive description of the intervention. Whereas a contract was in place between the private partner and the private companies (service providers), there was no direct contract between the university and the service providers. Therefore, responsibilities were unclear, yet some data collection activities depended on these service providers. These data collection activities may have been perceived as an additional task with no agreed upon compensation. Having no binding agreements in place and without a thorough understanding of the study resulted in missing or incomplete data.

Treatment group provisions fit the definition of a complex intervention. Because we were not able to fully follow the lessons learned by the Institute of Education Sciences (IES) (<http://ies.ed.gov/>) that state sequential steps are critical in researching interventions (beginning with development, pilot testing and refining), the intervention remained under developed and under refined as participants began enrolling into the study. The first handful of participants

was exposed to an intervention that was essentially, a moving target. Although we brought this up to the cross-site evaluators, they continued to reinforce the RCT design.

The project did not develop a written description of the intervention model that standardized and operationalized the treatment components, which would have enabled development of a logical chain of reasoning to link intermediate and long-term outcomes of interest. As a result, we were challenged with identifying appropriate measures relating to specific outcomes and impact.

Despite threats to evaluation integrity, we responded with flexibility keeping within the framework of rigor and relevance. The Joint Standards served as a foundation for our activities while the AEA Guiding Principles guided our behaviors. Efforts to establish our place within the organization and study structure took much time and energy. Given the circumstances, we believe there were valuable lessons worth sharing.

Discussion/Lessons Learned

We began this discourse describing the context of a community-based study consisting of a complex intervention within a public/private partnership situated alongside an independent evaluation employed by the same organization as the project. The mandated experimental design, often promoted as the gold standard compared to other designs, permitted a rigorous investigation of intervention effectiveness. This paper sheds light on many of the challenges and issues we faced throughout the evaluation process and offers a firsthand account of how the practice of evaluation can be unintentionally

compromised. Evaluator integrity was defined by describing how we chose to respond to the conditions placed before us.

The fields of nursing research and pharmaceutical studies portray collaborations between public and private entities with much written about successful endeavors. The World Bank has its own Independent Evaluation Group (IEG) that reports directly to the Bank's Board of Executive Directors and has a protocol of how it functions in relation to the Bank (http://siteresources.worldbank.org/EXTDIRGEN/Resources/dge_mandate_tor.pdf). However, few articles detail the complexities involved in these undertakings and more importantly how evaluators negotiate the boundaries of rigor and relevance while operating in these situations. In trying to design the evaluation plan and data collection activities, we were concerned with a re-occurring issue of the PPP implementing a service-oriented program versus an empirical intervention. These are two core disparate trains of thought surrounded by the larger challenges of the public/private partnership and the organizational placement of the evaluation. Because partners lacked experience conducting research or a comprehensive understanding of evaluation properties, our role was sometimes seen as divisive and adversarial and our responsibilities often misunderstood.

Stated succinctly by Shadish (2006), "The role of program evaluator is not the same as the role of program developer. Indeed, some evaluators argue that the two roles are incompatible because the developer is often biased toward wanting a positive evaluation of the program" (p. 3). Both partners needed constant reminders about the federal contract and

the expectations to conduct a meticulous evaluation. Our efforts to guide activities toward contract compliance yet appropriately support the PPP were often met with accusations of being non-collaborative.

Had there been time to thoroughly educate the staff about the research process and then introduce the evaluation requirements perhaps the challenges would have been easier to overcome. However, it is not uncommon for people to feel threatened by anyone looking at their practice, and evaluators may find resistance regardless of how responsive or sensitive their actions (Gaskill, et al, 2003).

It was critical that we kept the partners informed and educated about our goals and further avoid threats to research integrity. In addition, because evaluation and project staff were both employed in the same organization, what seemed like office politics often times became questions of ethics and integrity. In some instances, routine evaluation activities (i.e. stakeholder interviews, data confidentiality, sub-contracting services) were questioned by the project director in an effort to “supervise” our work. Relationships were damaged as a result of how the project and evaluation were organizationally administered with our efforts to uphold funding requirements and evaluation ethics.

Recommendations

In these times of limited resources, although public/private partnerships are necessary, economical and logical, they bring together two opposite frames of thought and processes. The rigorous research process required by the funders coupled with the fast track consumer

driven private business sector resulted in additional challenges for the evaluation of this project. The consolidation of resources by housing both project and evaluation teams in one organization fundamentally did not work. However, we are not convinced that it cannot work under optimal conditions. The challenges inherent in our case may no longer be unique as organizations leverage funds to gain the most bang for their buck. Evaluators too, must be proactive in how to maneuver around these environments while minimizing threats to integrity. We conclude this paper with a set of recommendations for organizations/institutions, evaluators, project developers/ implementers, and grantors.

Recommendations for Organizations/Institutions Conducting both the Implementation and Evaluation of the Same Project

We recommend that a protocol similar to the World Bank’s IEG protocol be implemented so that all employees are aware, knowledgeable and respectful of where an independent evaluation is placed within a company. Administering a dual activity grant necessitates a willingness to admit that there are competing objectives. However, it is essential that organizational leaders actively support the independent evaluation as a viable asset. Particular to our situation, project staff needed more education about the purposes of evaluation and how this particular evaluation should function within the organization.

When initiating any type of partnership, it is recommended that

institutional guidelines and policies be developed at the onset of the interactions to prevent some foreseeable problems, facilitate communication, and promote positive relationships (Walt, Brugha, & Haines, 2002; Barr, 2007). MOUs or MOAs are not sufficient to hold partners accountable for their responsibilities.

Recommendations for Evaluators

Evaluators should develop a set of operational guidelines that clearly define the roles and responsibilities of an independent evaluator and how the Joint Standards and Principle Guidelines apply to the evaluation activities. These guidelines go beyond a contract or scope of work. The operational guidelines would be used to educate project staff about the goals of the evaluation and thoroughly explain the design requirements. Related to this document is the need to develop trust among all parties coupled with time to nurture a professional relationship.

We recommend that evaluators think beyond common practice. Data collection activities can seem obtrusive and extraneous to those unfamiliar with evaluation practices. In many instances and when appropriate, we negotiated data collection activities such as using audio-tape recordings rather than videotape; minimizing the amount of intervention staff data collection requirements and integrating unplanned research questions into the evaluation design.

Recommendations for Project Developer/Implementers

Project staff, unfamiliar and/or uncomfortable with evaluation requirements should seek out general evaluation professional development

opportunities. Rather than viewing evaluation as a necessary evil (monitoring), we encourage the view of evaluation as being supportive and providing guidance. Organizations that develop and implement programs could provide “brown bag” luncheons that informally educate staff about evaluation basics.

Recommendations for Grantors

To avoid inconsistent lines of communication and authority, grantors should contract separately with independent evaluators and not allow project grantees to sub-contract or interpret the definition of “independence.” Organizations such as the Robert Wood Johnson Foundation elicit proposals for evaluations of funded projects and programs connected to the foundation. We endorse the recommendation that grantors should consider the merit and purpose of facilitating a multi-tiered evaluation structure for their activities (Rodi & Paget, 2007).

Through this experience, we found ourselves at times being micro-managed and at other times having no guidance. Accountability was often tossed between the federal funders and their contracted evaluators. We believe one reason this occurred is that the federal project officers had no experience in evaluation and often deferred to the multi-site evaluators. Thus, another recommendation for grantors is to ensure that project officers have a conceptual knowledge of evaluation and are competent in the basics of evaluation theories and activities.

Concluding Thoughts

After seeking counsel from our advisory panel and other evaluators at different venues, the overarching recommendation was to continue in the direction we were pursuing and consider this a valuable learning experience. As a result we have learned additional ways to think about research design, survey development and data collection. We have also sharpened our skills working with peers unaccustomed to research practices and interacting with members from the community, including private sector organizations.

The rigorous research process required by the funders coupled with the fast track consumer driven private business sector resulted in numerous challenges for the evaluation of this demonstration. While at times, frustrating the opportunity to learn and apply these learnings to other settings encouraged us. Furthermore, we believe that partnerships between public and private entities collaborating to implement rigorous research hold promise for future endeavors.

References

- American Evaluation Association. (2004). *Guiding Principles for Evaluators*. <http://www.eval.org/publications/GuidingPrinciplesPrintable.asp>. Retrieved May 2, 2010.
- Chaskin, R.J. 2003. The challenge of two-tiered evaluation in community initiatives. *Journal of Community Practice* 11(1), 61-83.
- Fitzpatrick, J. L., Sanders, J. R. and Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines*. New York: Pearson Education Inc.
- Gaskill, D., Morrison, P., Sanders, F., Forster, E., Edwards, H., Fleming, R. & McClure, S. (2003). University and industry partnerships: Lessons learned from collaborative research. *International Journal of Nursing Practice*, 9, 347-355.
- Gorman D. M. , Conde, E. (2007). Conflict of interest in the evaluation and dissemination of 'model' school-based drug and violence prevention programs. *Evaluation and Program Planning*, 30, 422-429.
- Medical Research Council. (2008). Developing and evaluating complex interventions: New guidance. *BMJ*, 337, a1655.
- Powell, W. & Koput, K. &.-D. (1996). Interorganizational collaboration and the locus of innovation: networks of learning in biotechnology. *Administrative Science Quarterly* , 41, 116-145.
- Rodi, M. S. & Paget, K. D. (2007). Where local and national evaluators meet: Unintended threats to ethical evaluation practice. *Evaluation and Program Planning*, 30(4), 416-421.
- Shadish, W. R. (2006). The common threads in program evaluation. *Preventing Chronic Disease*, 3(1), 1-5.
- Walt, G., Brugha, R., & Haines, A. (2002). Working with the private Sector: the need for institutional guidelines. *BMJ*, 432-435.