How Can Multi-Site Evaluations be Participatory?

FRANCES LAWRENZ AND DOUGLAS HUFFMAN

ABSTRACT

Multi-site evaluations are becoming increasingly common in federal funding portfolios. Although much thought has been given to multi-site evaluation, there has been little emphasis on how it might interact with participatory evaluation. Therefore, this paper reviews several National Science Foundation educational, multi-site evaluations for the purpose of examining the extent to which these evaluations are participatory. Based on this examination, the paper proposes a model for implementing multi-site, participatory evaluation.

INTRODUCTION

Over the past decade the Education Directorate of the National Science Foundation (NSF) has been increasing the proportion and amount of support it provides for large national initiatives, in contrast to the smaller grants it has made in the past. Simultaneously, all governmental agencies have become more and more concerned about evaluation (Government Performance and Results Act, 1993). NSF is faced with the difficult challenge of providing high-quality evaluation for all of its programs. This challenge is compounded by the fact that most of the programs are delivered in a multi-site format. For example, the Education Directorate funds multi-site programs such as Local Systemic Change (LSC), Collaboratives for Excellence in Teacher Preparation Program (CETP), the Centers for Learning and Teaching (CLT), and Advanced Technological Education (ATE). NSF's substantial new effort, the Mathematics and Science Partnerships (MSP), is also a multi-site program. These programs pose an especially interesting evaluation problem because each site within a program is in itself complex. The sites are not necessarily implementing similar procedures or materials and each often has a unique approach to solving the national issue addressed by the program. For example, LSC "sites" could be a single, large school district or a consortium of several districts, working at the

Frances Lawrenz • Professor and Chair, Department of Educational Psychology, College of Education and Human Development, 206 Burton Hall, 178 Pillsbury Drive SE, University of Minnesota, Minneapolis, MN 55455, USA; Tel: (1) 612-625-2046; E-mail: Lawrenz@umn.edu.

American Journal of Evaluation, Vol. 24, No. 4, 2003, pp. 471–482. All rights of reproduction in any form reserved.
ISSN: 1098-2140 © 2003 by American Evaluation Association. Published by Elsevier Inc. All rights reserved.

elementary school or high school level, and emphasizing mathematics or science. What ties the LSC sites together is a focus on teacher professional development and system-wide support to enhance student understanding of mathematics and science. Therefore, the evaluations of this and similar programs fit Sinacore and Turpin's (1991) notions of multi-site evaluation: They involve multiple sites and require cross-site evaluation activity. The purpose of this paper is to review the evaluations of these National Science Foundation educational, multi-site programs to consider the extent to which they are participatory. Based on this examination, the paper proposes a model for implementing multi-site, participatory evaluation.

Two volumes of the New Directions for Evaluation (NDE) series have discussed multi-site evaluations. The older volume, *Multisite Evaluations*, edited by Turpin and Sinacore, focuses on issues of multi-site evaluations irrespective of discipline. The chapters discuss the different benefits and challenges in conducting multi-site evaluations related to staffing, quality control, adaptation to local needs, generalizability, and statistical analyses. The more recent volume, *Conducting Multiple Site Evaluations in Real-World Settings*, edited by Herrell and Straw, focuses on the evaluation of substance abuse and mental health interventions. Straw and Herrell (2002) begin by expanding on the definition of multi-site evaluations provided by Sinacore and Turpin (1991). They point out that multi-site evaluations are different from cluster evaluations or multi-center clinical trials and that they are usually federally funded. In the concluding chapter, Leff and Mulkern (2002) suggest that two principles play an important role in multiple site evaluations: the science-based principle and the participatory principle. The science-based principle refers to the classic notions of research design where interventions are carefully tested against comparisons, and Leff and Mulkern discuss how multi-site evaluations can promote these design standards. The participatory principle implies that stakeholder groups should have meaningful input in all phases, such as designing the evaluation, defining outcomes, and selecting interventions.

Theoretically, infusing multi-site evaluations with the participatory evaluation approach would provide high-quality evaluations, by capitalizing on the capacity and experiences of the multiple sites and stakeholders. As suggested by Patton (1997a), stakeholder participation can enhance evaluation relevance, ownership, and utilization. However, in these multi-site program evaluations we need to examine what constitutes "participatory."

Generally, participatory evaluation means broadening bases for decision making and/or reallocating power in the production of evaluative knowledge (Cousins & Whitmore, 1998). This works very well when the object of the evaluation is a small, single site such as a school. In small cases, the teachers, parents, principal, students, school board members, district personnel, and perhaps evaluation specialist(s) and relevant external funder(s) can meet together to determine the evaluation questions, the data necessary to answer them, and how to gather and use the data. These people may all be familiar with each other and the important issues at the school. A participatory approach is feasible in that the various stakeholders can all contribute to decisions about how the evaluation should operate. In the large, multi-site evaluations of the type reviewed here, however, there are many different layers of stakeholders, there is wide diversity, and each site is often unfamiliar with the others. Furthermore, the sites were selected because of their success in a competitive grant process, not because they would facilitate the program evaluation in general, let alone a multi-site participatory evaluation. Each "site" is also a combination of sites, each with its own stakeholders, and is a complex, expensive project funded for many thousands to millions of dollars.

For example, each Collaborative for Excellence in Teacher Preparation (CETP) site involves the project leaders, evaluators, and several colleges and school districts, with their

attendant administrators, faculty/teachers, and students, as well as NSF and all of the other funding agencies involved in supporting the project. Obviously having all of the stakeholders from all of the sites interact in such a multi-site program is impossible. Some sort of representational system is necessary, with representatives of different communities of stakeholders or sites involved in the decision making. There could be different participatory sets, such as the stakeholders in each of the participating districts and colleges. Each set could send a representative to the management team which would include other types of stakeholders such as evaluators, the project leaders, and other funders. Each site management team could send a representative to the program evaluation team where decisions about what to evaluate and how would be made. For the purpose of this review, therefore, we use the extent to which individual projects or sites are involved in the program evaluation as a measure of the program evaluation's participatory nature.

As a consequence of this representational approach, the level of competency of the participants in a program level evaluation is often quite high. One of the valuable aspects of participatory evaluation is the development of evaluation capacity (Whitmore, 1998). Often, this capacity is developed among people who do not have prior evaluation skills, or evaluation capacity infused within an entity in terms of embedded evaluation processes where none existed before (Baizerman, Compton, & Stockdill, 2002). Multi-site participatory evaluations offer a unique opportunity for evaluation capacity-building at an advanced level. One component of this capacity-building is the opportunity that the interaction of evaluators from each of the sites provides for improving the quality of the overall evaluation (Leff & Mulkern, 2002).

Deciding on the measure of participation is not enough. The extent of involvement needs to be determined as well. Two ways to consider the extent of participation were proposed in Whitmore's (1998) NDE volume, *Understanding and Practicing Participatory Evaluation*. Cousins and Whitmore (1998) propose a three-dimensional formulization of collaborative inquiry. The three dimensions are: control of the evaluation process, stakeholder selection for participation, and depth of participation. Burke (1998) suggests that the process of participatory evaluation has a spiral design with key decision points. The decision points are (a) deciding to do it, (b) assembling the team, (c) making a plan, (d) collecting data, (e) synthesizing, (f) analyzing and verifying the data, (g) developing action plans for the future, and (h) controlling and using outcomes and reports.

In this section we have proposed project involvement in program evaluation as a measure of participation and referenced two procedures for considering the extent of participation. In the next section, we provide descriptions of the different program evaluations and consider project involvement in terms of Cousins and Whitmore's and Burke's ideas.

DESCRIPTION OF THE PROGRAM EVALUATIONS

In order to consider the aspects of participation in multi-site program evaluations, five different, large NSF program evaluations are described. These descriptions are based on our personal experience with the programs and their evaluations.

The Centers for Learning and Teaching (CLT) program is designed to produce leaders in science and mathematics education through the development of centers for research on learning and teaching. The CLT program evaluation plan includes a set of surveys for various participants as well as yearly site visits to all projects. An externally-funded evaluation, it employs its own instruments and collects its own data independently from the funded centers. The multi-site

evaluation is directly tied to the needs and questions of NSF but is not necessarily tied to the needs and ideas of the CLTs themselves. A logic model for the CLT program was developed by the program evaluation team in conjunction with NSF, and evaluation survey instruments and site visit processes were developed based on that model. Reports will be supplied directly to NSF and not necessarily shared with the projects. In terms of Cousins and Whitmore's (1998) dimensions, the control is completely centralized. The selection of participants or stakeholders is completely in the hands of the centralized evaluation. Participation at the sites is limited to responding to the instruments. In terms of Burke's key decision points, the first decision precluded all of the other decision opportunities. The decision was made to not include the project sites in the development and conduct of the program evaluation.

The Advanced Technological Education program (ATE) is designed to increase the number and quality of technicians in the nation. It is focused on community colleges and includes activities such as collaboration, materials development, program improvement, and professional development. The ATE program evaluation has consisted of a web-based survey of all the projects and a series of site visits to 13 project sites. The ATE program evaluation is slightly different than the CLT program evaluation. The level of information asked for in the web-based survey is already expected to be aggregated to some degree by the projects' principal investigators; thus, the evaluator does not directly acquire information from the various individuals involved in the project. Additionally, the ATE evaluation has two advisory groups with expertise and experience with the ATE program to help guide the evaluation process. These advisory groups were used to help construct the evaluation instruments and to conduct the site visits. Additionally, the web-based survey questions are shared with the projects and periodically revised based on project input. Furthermore, the results of the survey are posted on the web for use by the projects. As with the CLT evaluation, the site visit protocols were developed by the program evaluation team; however, the reports from each of the site visits were shared with the site and not shared with NSF. In this regard, there is a slightly broader selection of stakeholders than in the CLT evaluation. However, the control is completely with the program evaluation team, and the depth of participation is minimal. Considering Burke's decision points, the ATE provides the evaluation information to the projects, and the evaluation is receptive of feedback, so it has made participation-oriented decisions in terms of considering the data and in using outcomes.

The Local Systemic Change (LSC) program is designed to improve K-12 science and mathematics education through a focus on school district-wide change and teacher professional development. Each funded project is required to gather specific information using pre-designed evaluation instruments. Each project may also add its own evaluation components. Specified procedures are required of every project, and the projects did not have any input into developing the evaluation procedures. However, those projects in existence at the very beginning of the evaluation effort were consulted about the adequacy of the evaluation procedures. Additionally, the evaluation effort has been evolving with changes based in part on comments from participating projects. In terms of control, the evaluation is handled by the program evaluation team. In terms of depth of participation, the individual projects gather their own information and make their own judgments about it, both of which are used in the ensuing program evaluation. All projects are included in the program evaluation. In terms of Burke's decision points, the first few decisions resulted in project evaluators being substantive but not powerful or decision-making members of the program evaluation team. Project evaluators gather the data and analyze and verify it to some degree. The synthesis of the data, the development of future plans, and the use of reports is completely in the hands of the program evaluation team.

The Collaboratives for Excellence in Teacher Preparation (CETP) program is designed to improve mathematics and science teacher preparation through the improvement of undergraduate science and mathematics courses and education courses. This improvement involves the collaboration of education and science and mathematics departments within a given institution, as well as collaboration with other institutions including community colleges and K-12 schools. The CETP program evaluation went through an evolutionary process, initially with each site collecting the information it felt was relevant, and moving to a plan whereby all sites collect some similar data using centrally-developed instruments. The procedures were developed by the participants, and all the projects had input into the instrument development. Projects can decide which data they wish to provide or not provide. There is control in the sense of building consensus and developing instruments based on that consensus. The program evaluation or core team provides leadership, a communication hub, instruments, data analysis, and provision services and incentives for collecting core data. Selection of stakeholders began with everyone in the program, although each project can decide whether or not to participate. Depth of participation varies according to the value each project sees in the core data. In terms of Burke's key decision points, there were decisions to be participatory at each of the key points.

Mathematics and Science Partnerships (MSP) program is one of the newest NSF initiatives. It is designed to improve student understanding of mathematics and science through the development of partnerships of various institutions such as museums, colleges, school districts, etc. There is no multi-site program evaluation yet in place, but several research, evaluation, and technical assistance projects (RETAs) were funded to complement the MSPs. One of the RETAs is providing technical assistance to selected MSPs with the goal of developing a model for a MSP program evaluation. In terms of Burke's key decision points, there were no decisions as of yet to be participatory in a multi-site evaluation. For now, each project in the MSP program functions independently and will turn in its own individual evaluation report to NSF.

DEGREE OF PARTICIPATION IN DECISION MAKING

Analysis of the above examples suggests that a continuum of participation exists from no participation to complete participation of the sites in making decisions in the multi-site program evaluation. The examples represent different points along the continuum. Combining the ideas of Cousins and Whitmore (1998) and Burke (1998) with the multi-site examples resulted in four types of decision making for comparing the degree to which individual projects are involved in multi-site program evaluations. These four include sites making decisions about: (1) the type of evaluation information collected, such as defining questions and instruments; (2) whether or not to participate; (3) what data to provide; and (4) how to use the evaluation information. The types of participation in decision making for each of the examples are outlined in Table 1 and explained below.

At present, the MSP evaluation represents the no participation end of the continuum because the sites do not participate in a multi-site program evaluation at all. Any multi-site program evaluation of this program to date would result from a pooling of the individual evaluation reports. The next step on the continuum would be the CLT program evaluation. This evaluation is completely exterior to the sites, requires all sites to participate and provide all data, and does not allow any input from the sites in terms of how the evaluation information is used. Next are the ATE and LSC evaluations, which are centrally prescribed to the sites.

TABLE 1.
Participation of Projects in Program Evaluation Decision Making

	Information to be Collected	*Participation in Program Evaluation*	*Provision of Data*	*Use of Evaluation Information*
MSP	Not needed yet	Not needed yet	Not needed yet	Not needed yet
CLT	No input	Required	Required from individuals	None
ATE	Advisory committees	Required	Expected from site management	Advisory committees
LSC	Advisory committees	Required	Required from site evaluators	Advisory committees
CETP	Consensus	Voluntary	Voluntary	Partial control

These evaluations have some input from advisory committees on information to be collected and how the information should be used. The ATE expects data from project management (sometimes from evaluators), while the LSC requires data from the project evaluators. Closer toward the full participation of sites end of the continuum is the CETP evaluation, in which the evaluation information and use is determined through consensus and the provision of data and participation in the program evaluation are voluntary.

EFFECTS OF PROJECT PARTICIPATION

How can project participation in program evaluation decisions affect the program evaluation? In other words, in what ways can participation contribute to the overall quality of the evaluation? We suggest the following four specific dimensions of quality of an evaluation: (1) objectivity, (2) design of the evaluation effort, (3) relationship to site goals and context and (4) motivation to provide data. Objectivity refers to the degree to which the evaluation effort could be viewed as being conducted in an impartial fashion, as exemplified in the "scientific" program evaluation referred to by Feuer, Towne, and Shavelson (2002) or the gold standard referred to by Straw and Herrell (2002). Design of the evaluation effort refers to the likelihood of the program evaluation design being the best one possible as suggested by Leff and Mulkern (2002). Relationship to the site goals and context refers to the amount that the program evaluation effort is directly connected to what is going on at the site, an advantage stressed by Patton (1997a) in terms of increasing use of the results. Motivation to provide data refers to the connectedness the people providing the data feel with the program evaluation effort (e.g., "Do respondents feel this is important to me, because I think these data are relevant and answer important questions?"). People who are motivated to provide the data generally provide more complete and accurate information (King, 1998).

Project participation could be viewed as contributing negatively or positively to the objectivity of a program evaluation. This was the subject of a debate spearheaded by Stufflebeam, Scriven, Patton and Fetterman (Fetterman, 1997; Patton, 1997b; Scriven, 1997; Stufflebeam, 1994) from within their evaluation models. The debate was over whether or not people within a project could be objective. The "scientific" notion of evaluation (Feuer et al., 2002; Herrell & Straw, 2002) suggests that the evaluation be controlled by someone exterior so they are not biased in favor of the program. In the case of the CLT or ATE evaluations, the program

evaluators were working directly for the funder, NSF, with very little relationship to the projects and hence might be viewed as objective. However, objectivity can also be supplied by the use of experimental and comparison groups as suggested by Herrell and Straw (2002), or through the use of standardized, quantitative data collection devices (Nightingale & Rossman, 1994). Given that the site projects discussed here were funded through a competitive process and not for the purposes of conducting a "scientific" evaluation, their participation and agreement would be necessary to produce an experimental design.

Project participation could also contribute in a positive or negative way to the design of the evaluation. One possibility could be that because in a participatory evaluation more people are involved, the evaluation design would be of higher quality. For example, in the CETP evaluation the design improved through discussion with all of the project evaluators. However, it also takes a great deal of time to achieve consensus (King, 1998) and that might mean missing critical baseline data. Additionally, as suggested by Leff and Mulkern (2002), the methods chosen in a participatory evaluation may be driven by the "capacities of the least capable sites" (p. 97). Another possibility could be that the people selected to conduct exterior program evaluations are of higher capacity than those selected for project evaluation because of the complexity of program evaluation and, therefore, their designs would be superior. For example, the CLT, ATE and LSC evaluations were all conducted by well established centers in evaluation. On the other hand, given the complex nature of the large multi-site projects in all of the examples, their site-based evaluators have to be highly capable as well. It seems possible, therefore, that having project participation in the types of multi-site programs described here could result in a better evaluation plan than having no participation.

Although it is possible to have a program evaluation without any project participation that is directly related to the projects' goals and contexts, project participation is likely to increase the relationship between the program evaluation and the projects' goals. If the program evaluation is independent from the projects, it may fail to capture the unique aspects of the local context. On the other hand, consensus about what is appropriate in a program evaluation, such as was obtained in the CETP evaluation, does not guarantee that the program evaluation will be directly related to project goals. The CLT evaluation showed that lack of participation may result in apprehension on the part of the sites about just what is going to be evaluated. Additionally, the CLT projects were somewhat reluctant to provide the required information because they felt like they were in a game where they were unsure of the rules or the goals. Projects may also feel constrained in their own evaluation efforts, because they want to keep the response burden low and, therefore, do not want to duplicate questions, as was true in the ATE evaluation. Because the LSC evaluation requires participation of local evaluators, it appears to result in less apprehension and resentment. In some projects, however, the required components of the program evaluation results in constriction or lack of creativity in the local evaluation and focus it in areas that may not be the project's major concerns. This can be true even though the projects are part of the overall program and ostensibly reflect the program goals that the LSC program evaluation embodies.

The fourth dimension of effect of project participation in program evaluation is motivation to provide data. Because participation in multi-site program evaluation is necessarily representative in nature, this typical effect of participatory evaluation may be diluted. Motivation of actual data providers (e.g., teachers from a school) is dependent on the commitment and persuasive ability of their representative, who in turn may be quite removed from actual participation in evaluation decision making. Therefore, it is possible that the consensus obtained at the program level may not be supported at the base levels of the project. Additionally, as the

program evaluation continues over time the representatives to the program team may change or new projects may be added and the opinions of these new representatives may differ from the original ones. This type of evolution was shown in both the LSC and CETP evaluations. Providing the opportunity for project participation and control implies the opportunity for non-participation and consequently for incomplete or inadequate program level data. As the CETP evaluation has shown, nonparticipation can result despite substantial incentives.

A MODEL FOR PARTICIPATORY MULTIPLE SITE EVALUATIONS

Based on the discussion and examples above, we suggest that the ideal multi-site program evaluation model should be objective, have the mandate of the funder, provide the opportunity for site-based stakeholders to collect and interpret data, have the opportunity for sites to collaborate and develop evaluation questions and processes, and possess the ability to combine unique data from different sources. A model like this would address the problems associated with many multi-site evaluations, such as substantial variability in implementation, little information on actual practice, lack of common and appropriate outcome measures, and the need to synthesize results for evaluating large-scale multi-site educational reform programs (Hamilton et al., 2003). This does not imply a one-size-fits-all model for all multi-site program evaluations. Our intention is simply to suggest a model we believe could be useful in cases such as the examples provided, where the program evaluation collects data across sites, each site is large and complex itself, the sites are funded independently, and each site has responsibility for conducting evaluation.

We call this new approach *negotiated centralized evaluation*, because it is negotiated and is binding on the parties as in a treaty. It has an interior element in that each of the sites has its own goals, activities, stakeholders and evaluators, and it has an exterior element in terms of a central evaluation team which is not involved in evaluating any of the individual sites. The negotiated centralized evaluation model has three different stages, beginning with the creation of local, site-specific evaluation, and moving towards the negotiation of a centralized evaluation effort. The three stages in the model are: (1) creating the local evaluations, (2) creating the central evaluation team, and (3) negotiating and collaborating on the participatory multi-site evaluation.

Stage I—Creating the Local Evaluations

The first stage of negotiated centralized evaluation is the creation of the local evaluations. The local evaluations of each site form the foundation upon which the centralized evaluation is based. The individual evaluations of each site are critical to the ultimate success of the centralized evaluation and therefore, it is recommended that individual sites have the time and resources to plan and conduct an evaluation that fits their context and begins to answer the questions that are most important to their situation. During this stage, the individual sites should focus on developing evaluation questions that are aligned with their site goals and on beginning to develop instrumentation and measurement techniques that can begin to answer their evaluation questions.

One difficulty for the local projects in a negotiated centralized program evaluation is the need to alter initial evaluation procedures to fit future negotiated ones. This was the case in the CETP program evaluation which began several years after the program started funding projects.

Ideally, the individual project evaluations serve as the starting point for the future negotiation of the central program evaluation. The sites should experiment with different procedures and potential outcomes. These grounded, data-rich experiences would eventually be followed by the group's finalization of what should be considered central and how it should be measured. This would contribute to the values base and the adequacy of the evaluation design as discussed by Leff and Mulkern (2002).

In a sense, the local evaluations play a key role in the establishment of a baseline of program data. Collection of baseline data is a particularly difficult issue if the multi-site program evaluation is to be delayed because the data are not always easily merged. If the individual projects collect very different types of data, or if the data they collect is limited with respect to constructs which subsequently emerge as important, the program evaluation may not be successful. On the other hand, if the evaluation focuses on recurring cohorts the negotiated baseline data could be collected on the next cohort. It might also be possible to incorporate pre-existing data (e.g., test scores, school records) after the fact. What is likely is that the central program evaluation would eventually incorporate some of the baseline data proposed by each of the individual projects. Also ideally, the local projects would retain their original funding for project evaluation; then, with the central evaluation collecting "core" data, the project would be able to spend more of its money for unique or complementary evaluation efforts. This is presently the case in the CETP evaluation.

Stage II—Creating the Central Evaluation Team

Once the local evaluations efforts are established, the second stage involves the creation of the central evaluation team. We envision a small planning grant process whereby potential central evaluation leaders would propose how they would involve the projects in negotiation of the program evaluation. The central evaluation leader and the projects' representatives would produce a proposal for the program evaluation. This proposal could be required in a pre-specified amount of time at a pre-specified level of effort. The proposal planners would also interact closely with the funder. When completed, the proposal would be considered by the funder, probably through a peer review, and could be modified through the contracting process to fit the political constraints of the situation. This process is somewhat similar to the approach taken in the MSPs, where there is, at least presently, no multi-site program evaluation but several smaller evaluation grants. The proposal would outline the responsibilities of all parties in completing the program evaluation and what the consequences of not participating would be. Once funded, these responsibilities would be binding on all. This would mean that all projects presently in the program and all new projects would have to participate in the negotiated multi-site program evaluation. This is somewhat similar to the procedures in the LSC program evaluation, which was funded and then used advisory committees to develop instruments; however, in the case of negotiated centralized evaluation there would be more opportunity for initial negotiation and more flexibility regarding the details of the program evaluation. In contrast to the LSC evaluation, a negotiated centralized evaluation proposal does not imply that the program evaluation would be rigid or constant. Different types of projects within the program might participate differently or some might move in and out of the program evaluation as they began or completed activities. The program evaluation would be tailored to the needs of the projects and of the funder, rather than to a mandated central authority.

Stage III—Negotiation and Collaboration on the Participatory Multi-Site Evaluation

Ideally, a negotiated centralized evaluation would evolve slowly, beginning with the careful development of relationships (Cook, Carey, Razzano, Burke, & Blyler, 2002). In this stage, the focus should be on bridging the gap between the local evaluations at each site, on the one hand, and a broad central evaluation that can answer questions that cut across the sites, on the other. Gaining the support and cooperation of individual sites for the central evaluation is one of the first steps in this stage. Because collaboration will not occur on its own, it would need to be developed and nurtured through frequent communication and other trust building activities. As King (1998) suggests, leaders and nurturers are necessary in a participatory evaluation. Because not all sites participate in the same way and opinions about what should be done may vary, time for consensus building is crucial (Mowbray & Herman, 1991). A good way to support consensus building is to have the sites participate in instrument and procedure refinement and interpretation of the collected data, similar to the "invisible college" idea suggested by Boruch (1991) where everyone is intellectually engaged in developing meaning. The importance of collaborative instrument development is highlighted by Cook et al. (2002) in their discussion of a multi-site employment intervention demonstration program. Collaboration allows for the pooling of ideas and synergistic effects, although care must be taken so that consensus does not devolve to the lowest common denominator (Leff & Mulkern, 2002).

It must also be kept in mind that intellectual willingness may not translate into collaborative behavior unless conditions support the behavior. One way to support collaborative behavior is to build evaluation capacity at each of the sites which is designed to support the negotiated centralized evaluation (Baizerman et al., 2002). Additionally, because incentives to sites may not be sufficient to support collaboration, sites should have to agree to participate before receiving funding. As Burke (1998) suggests, participatory evaluations should not be attempted unless all stakeholders are in agreement with the process. Leff and Mulkern (2002) extend this to the notion that the funders must pay more attention to when and how to conduct multi-site evaluations and to when and how to involve non-evaluator stakeholders. Consequently, the funders and the sites would have to be in complete agreement with using the negotiated approach and conducting the program evaluation in accordance with the developed evaluation design. All components of the power structures would have to agree to participate in the consensus building or negotiation and be bound by the results. As new projects are funded, they could be folded into the negotiation process although they would not have as much input as the beginning sites. However, new projects would have the advantage of knowing what would be provided by the program evaluation and they could tailor the project evaluation in their proposals to complement it.

Having a central team can be an advantage in a negotiated centralized evaluation. It can help to provide objectivity because the central team is not involved in any site-specific evaluation. The central team could also be responsible for collecting or analyzing some data themselves, for example, collecting and rating artifacts such as lesson plans, which would further extend the objectivity. This objectivity could also extend to paying careful attention to quality control through the standardization of data collection and the development of valid and reliable evaluation instruments as described by Sinacore and Turpin (1991). Centralization can also be advantageous because sites can share the workload. By agreeing on central questions and sharing the work involved in developing instruments and processes, individual sites can be freer to concentrate on the unique aspects of their approaches. Furthermore, the central team can provide data collection, input and analysis services for all, thereby providing an economy

of scale. Additionally the central team can present a more consolidated front in negotiations with the funder and serve as an intermediary between the individual sites and the funder. The central team can also facilitate the sharing of expertise across the sites (Leff & Mulkern, 2002).

SUMMARY

To assist in producing the type of program evaluations necessary for large-scale, multi-site Federal programs, we have proposed a different model of multi-site program evaluation. It is a model of representative participatory evaluation which values the contributions of each site, while at the same time encouraging collaboration. This approach offers several advantages, but must be implemented carefully to capitalize on them. In the negotiated centralized model, the multi-site evaluation plan would evolve from investigations at the sites. As a result, the instruments and processes will be grounded in the reality of the program as it is implemented, while at the same time the multi-site evaluation should produce the kind of external, objective examination of programs required by the Federal government. If the accountability movement manages to overshadow individual project evaluation in the name of "scientific" evaluation, we will have lost critical information about unique program impacts. It is imperative that new models of evaluation be developed to help meet the requirements of objective, scientific evaluation while, at the same time, valuing and incorporating local evaluation efforts. The negotiated centralized evaluation model described here is one example of how to meet this need.

ACKNOWLEDGMENTS

This material was partially based upon work supported by the National Science Foundation grants DUE-9908902 and REC-0135385.

REFERENCES

Baizerman, M., Compton, D., & Stockdill, S. (2002). New directions for ECB. In D. Compton, M. Baizerman, & S. Stockdill (Eds.), *The art, craft, and science of evaluation capacity building* (pp. 109–120). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation, 93*.

Boruch, R. (1991). Sharing confidential and sensitive data. In J. E. Sieber (Ed.), *Sharing social science data: Advantages and challenges*. Thousand Oaks, CA: Sage.

Burke, B. (1998). Evaluating for a change: Reflections on participatory methodology. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 43–56). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation, 80*.

Cook, J., Carey, M., Razzano, L., Burke, J., & Blyler, C. (2002). The pioneer: The employment intervention demonstration program. In J. Herrell & R. Straw (Eds.), *Conducting multiple site evaluations in real-world settings* (pp. 31–44). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation, 94*.

Cousins, J., & Whitmore, E. (1998). Framing participatory evaluation. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 5–24). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation, 80*.

Fetterman, D. (1997). Empowerment evaluation: A response to Patton and Scriven. *Evaluation Practice, 18*, 253–266.

Feuer, M., Towne, L., & Shavelson, R. (2002). Scientific culture and educational research. *Educational Researcher*, *31*(8), 4–14.

Hamilton, L., McCaffrey, D., Stecher, B., Klein, S., Bobyn, A., & Bugliari, D. (2003). Studying large-scale reforms of instructional practice: An example from mathematics and science. *Educational Evaluation and Policy Analysis*, *25*(1), 1–29.

King, J. (1998). Making sense of participatory evaluation practice. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 57–68). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation*, *80*.

Leff, H., & Mulkern, V. (2002). Lessons learned about science and participation from multisite evaluations. In J. Herrell & R. Straw (Eds.), *Conducting multiple site evaluations in real-world settings* (pp. 89–100). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation*, *94*.

Mowbray, C., & Herman, S. (1991). Using multiple sites in mental health evaluations: Focus on program theory and implementation issues. In R. Turpin & J. Sinacore (Eds.), *Multisite evaluations* (pp. 45–58). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation*, *50*.

Nightingale D., & Rossman S. (1994). Managing field data collection from start to finish. In J. Wholey, H. Hatry, & K. Newcomer (Eds.), *Handbook of practical program evaluation* (pp. 350–373). San Francisco, CA: Jossey-Bass.

Patton, M. (1997a). *Utilization-focused evaluation* (3rd ed.). Thousand Oaks, CA: Sage.

Patton, M. (1997b). Of vacuum cleaners and toolboxes: A response to Fetterman's response. *Evaluation Practice*, *18*, 267–270.

Scriven, M. (1997). Comments on Fetterman's response. *Evaluation Practice*, *18*, 271–272.

Sinacore, J., & Turpin, R. (1991). Multiple sites in evaluation research: A survey of organizational and methodological issues. In R. Turpin & J. Sinacore (Eds.), *Multisite evaluations* (pp. 5–18). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation*, *50*.

Straw, R., & Herrell, J. (2002). A framework for understanding and improving multi-site evaluations. In J. Herrell & R. Straw (Eds.), *Conducting multiple site evaluations in real-world settings* (pp. 5–16). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation*, *94*.

Stufflebeam, D. (1994). Empowerment evaluation, objectivist evaluation and evaluation standards: Where the future of evaluation should not go and where it needs to go. *Evaluation Practice*, *15*, 321–338.

Whitmore, E. (1998). Final commentary. In E. Whitmore (Ed.), *Understanding and practicing participatory evaluation* (pp. 95–99). San Francisco, CA: Jossey-Bass. *New Directions for Evaluation*, *80*.