INTRODUCTION

A recent New Directions for Evaluation “Conducting Multiple Site Evaluation in Real-World Settings” focused on the evaluation of substance abuse and mental health interventions. 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. Here we extend the participatory principle to the field of teacher preparation using the example of the core evaluation of the Collaboratives for Excellence in Teacher Preparation. The recent passage of the “No Child Left Behind” Act and the U.S. Department of Education’s “Before it's too late” report (2000) have heightened awareness of the need for well prepared teachers, especially in science and mathematics. Improving the preparation of science and mathematics teachers has long been the province of the National Science Foundation and one of their more recent efforts has been the Collaboratives for Excellence in Teacher Preparation Program (CETP).

According to the NSF, the CETP program promotes comprehensive change in the undergraduate education of future teachers by supporting cooperative, multiyear efforts to increase substantially the quality and number of teachers well-prepared in science and mathematics, especially members of traditionally underrepresented groups. Nineteen System-Wide CETPs were funded from 1993 through 2000 when the program stopped receiving new proposals. Although each CETP was unique, there were some similar characteristics. All engaged faculty from the sciences, mathematics and education. All included several institutions of higher education (including community colleges) in a particular geographic area, e.g., a state. All included mechanisms for improving undergraduate education in the sciences and mathematics. All had some sort of relationship with schools within the geographic area. All offered scholarships to students from underrepresented groups. The ranges within these were broad however e.g., from 3 institutions to 13 or more, and all had unique elements like liaisons with Tribal Colleges or industrial/science internship placements, etc.

Each collaborative was, or is anticipated to be, funded for five years to accomplish the teacher preparation goals and then for an additional three years to more closely examine the effects of the Collaborative. Evaluation of programs plays a prominent role at NSF. The Government Performance and Results Act (GPRA) passed in 1993 requires that Federal agencies determine measurable goals for all of their program activities. Agencies must then measure their performance against these goals and report progress to Congress as part of the annual budget submission.

THE EVOLUTION OF THE CETP CORE EVALUATION

The first significant event was a meeting of the different CETP evaluators in February, 1998 just as the first three funded CETPs were eligible for the second phase of funding. It resulted in a report “Corridors for Collaboration” which described potential corridors for enhancing the evaluation of the CETP program through increased collaboration of
evaluators. The meeting highlighted thoughtful potential solutions to many challenges faced in the evaluation of these comprehensive programs and the willingness of the evaluators to work together. The meeting was facilitated by a “neutral” party who had worked as a consultant with several of the CETPs. Using this person helped control competition among the CETPs but also left the group without internal leadership once the meeting was over.

NSF sponsored a second meeting in November, 1998 to promote the initial collaboration and to focus on defining centralized evaluation needs. This second meeting highlighted the need for continuous external promotion of collaboration and stakeholder involvement. The reality of trying to meet the needs of the locally specified evaluation plans precluded any real collaboration. Once again the evaluators and their projects were very positive toward collaborating on evaluation efforts. There was talk of sharing evaluation instruments and processes but no one emerged as a leader willing to take responsibility for this to happen. The NSF wanted the evaluation to emerge from the projects and be a voluntary effort rather than a required, contracted evaluation. The evaluators, however, felt that they had no time for the extra effort involved in collaboration.

Consequently NSF funded a Core Evaluation Project managed by the formerly neutral party. This provided consistency and expertise. The ongoing CETP Core Evaluation Project proposed spending the first year (1999-2000): 1) achieving consensus on what evaluation questions the core evaluation should address, 2) evaluation instrument development tied to the core evaluation questions, and finally, 3) pilot testing to gather data to improve the developed instruments. The year of 2000-2001 was set aside for the collection and analysis of field test data and for the development of draft reports utilizing the collected data. The data and the reports would be used to verify that the type and amounts of data collection could meet the needs of the CETPs and NSF. The third year, 2001-2002, was intended for full data collection and dissemination. The philosophy of the Project was one of collaboration with the CETPs; the instrument development was viewed as a joint endeavor with the CETP Core Project providing technical assistance to the individual CETPs.

This project was funded as a grant and therefore the individual CETPs were not required to participate in the Core. The Core therefore had no “stick” and worked to provide “carrots” to encourage participation. Also the Core had to work with all of the CETPs; no selection of sites was allowed.

The Core was committed to collaboration and scheduled a meeting for the CETPs soon after receiving funding. The Core paid consulting time and travel. Before the meeting instruments from all of the CETPs were gathered and categorized and sent out to the CETPs who had volunteered to work on the sets of questions proposed at the second meeting. The intent was to develop sets of potential instruments or items that could be used to answer centralized evaluation questions. In October 1999 the “mature” CETPs (Those with 3 or more years of funding) met to finalize evaluation questions for the Core broader and more detailed questions were specified. The broad questions were:

Ia. How supportive of STEM reform education policies and procedures are the
The primary research questions focused on:

1. How well do CETP teachers demonstrate the knowledge and skill espoused by the standards?
2. How do CETP teachers and the classroom they create differ from non-CETP teachers and the classrooms they create?
3. What outcomes have the participating higher education institutions, their faculty or the CETP contributed to in the K-12 schools?
4. Are students learning what is expected in the STEM standards?
5. Are there differences in student outcomes for CETP and non-CETP teachers?

Several instruments were created to address these questions. The CETP evaluators provided good critiques when they had the time, but the bulk of the work was accomplished by the Core. Ten different sources of data were identified: Higher educational institution deans/department chairs, faculty, students, and classrooms, K-12 principals, teachers, students and classrooms, NSF scholars and the CETP PI/evaluators. All of the items on the instruments were matched to the evaluation questions and sources. Instruments included web based surveys for the higher education deans and faculty members and the PIs; paper and pencil surveys for college students and K-12 principals, teachers and students; classroom observation protocols with web based data entry capacity and rubrics for scoring two classroom artifacts, activity sheets and assessments. All of these were pilot tested and revised based on data from samples of each respondent group. A listserve was created to promote dissemination of information and discussion. All of the instruments and a cross walk to the evaluation questions are posted on the web site www.education.umn.edu/CAREI/CETP.

A second meeting of the mature CETPs was held in fall 2000 for them to go over the instruments and make suggested changes. They were paid by the Core to attend and for any substantial prior work on revisions. After the meeting the instruments were finalized and made available to the CETPS to collect data. The Core had money available to pay the CETPs for the extra work involved in collecting the Core data. As an additional incentive to participate, the Core agreed to enter, clean and specify all data from a CETP and return it to them as an SPSS system file so that they could conduct their own analyses with the data. A comprehensive sampling plan and a list of responsibilities for the CETPs and the Core were outlined and data collection began.

Although all CETPs attended the meeting, the money for revision work had not enticed many of the evaluators to work on the instruments so that work fell mostly to the Core. Although all of the CETPs had agreed on the evaluation questions and everything was discussed and finalized, the agreement was not always enthusiastic. Some CETPs felt their existing instruments were better matched to their CETP than the Core instruments and some felt that the Core instruments should be modified more. Although consensus didn’t result in the methods chosen “driven by the capacities of the least capable sites” as suggested by Straw and Herrell (2002), it did result in compromises. There was a good
espirit-de-corps built up within the group and strong positive interpersonal relationships among the CETP evaluators and the Core PI and staff.

Simultaneously NSF suggested that the Core work with the less mature CETPs as well so a meeting was held for them at the Core’s expense in January 2001. This group was very enthusiastic about the instruments and many planned to use them in their evaluations even though they were not part of this field test year data collection.

The data collected in the field test was very sporadic; only some of the specified data were collected and from only some of the CETPs. One CETP might have K-12 student data but no principal data. One CETP might have all of the higher education data while another CETP had none. Few CETPs used the money available from the Core to help collect data. They said it was difficult to find qualified people to hire to handle data collection and many evaluators felt it was just easier to do it themselves. They did not really have the time and therefore no data were collected. Also some CETPs felt they did not want to collect certain types of data, despite their agreement that the data and instruments were appropriate for the Core. They saw their circumstances as unique. However, as would be expected from the strong interpersonal relationships, the CETPs were very apologetic to the Core for not collecting the data. The listserve was used by the Core to send information but was not used by the CETPs. The CETPs preferred to correspond personally with the Core PI and staff which helped to cement the interpersonal relationships. The Core suggested to NSF that the Core project be discontinued on the basis of the poor response but NSF encouraged the Core to try another year. NSF sent letters to all of the CETP PIs stating that NSF saw the Core as a valuable effort.

The Core tried to think of more ways to provide incentives for participation in the data collection. The Core sponsored another meeting, this time with all of the CETPs. The focus of the meeting was on what each CETP could do to provide data to the Core. “Good” data collectors from the field test year were asked to give tips to others. The pattern of data collection (well lack thereof) was shown to all and the futility of trying to make statements about program success without a complete data set was discussed. The preliminary results were shared and discussed to promote ownership in the interpretation and hopefully interest in collecting data. Instruments revised by the Core based on the field test data were discussed and finalized. The Core reiterated its ability to provide money to help the CETPs collect the Core data. All of the CETPs (mature and less mature) were included in the data collection. Everyone left the meeting enthusiastic about collecting data. The Core sent several reminders out during the year via the list serve and to individuals. NSF also changed its requirements for the three years of additional funding for the CETPs to include a question which asked how the project to be funded would work with the Core.

The data collected in 2001–2 were more comprehensive than the prior year but still inconsistent across CETPs and types of data. Some of the CETPs used the money available to help collect data but most did not. Despite the lack of data, the CETPs expressed the belief that the Core was worth while and that they wanted to participate.
The NSF supported another year of Core data collection.

**DISCUSSION**

This multi-site core evaluation highlights many of the complex issues that are inherent in a process that attempts to bridge the gap between centralized and individualized evaluation. While it is clear that collaboration and participation must be carefully supported and nurtured, as suggested by Herrel and Straw (2002), we argue that this is not enough. There are certain inherent characteristics related to the tension between centralized and individualized evaluation that must be considered to create participation in a core evaluation. Likewise, merely providing incentives to participate is not necessarily sufficient. The Core provided several incentives; interpersonal attachments and collegiality, opportunity for discussion and resolution of problems, efficiency of data collection through the use of pre-prepared instruments, return of collected data in a easily used system file, approval of the funding agency, and money to hire staff to help local sites collect data. Despite all of these incentives, the individual sites did not always participate.

This is not a criticism of individual sites or evaluators, but rather recognition of an important aspect of human behavior. The success of a core evaluation hinges upon the ability of the core evaluators to understand and consider group versus individualized behavior and a design that can capitalize on the strengths of each. While the individual sites believed in the core evaluation, wanted the core evaluation, helped to create the grassroots effort that created the core, in the end when it came time to collect data, they put individual needs first. The group actually only functions because it meets the needs of each individual, not vice versa.

An ironic curve in the road was that in some ways the CETP core was a victim of its own success. The grassroots movement for a core evaluation grew because CETP’s were too complex to evaluate within the allocated timeframes. However, NSF needed comprehensive project evaluations to justify its expenditures on the CETP program and therefore pushed the evaluators for even more information. Once established, the core evaluation served as an intermediary between the sites and NSF helping to moderate demands. The provision of the core evaluation questions left each site freer to concentrate on the unique aspects of its approach and to be less concerned about justifying the CETP program overall.

The development of data collection instruments and procedures provided diverse pathways. While the instruments were an excellent resource, they were not flexible enough to allow a truly participatory approach. They had to be consistent over time and across sites. Although all the sites viewed the core as good, as is often the case in consensus building, many felt the instruments and procedures were not exactly right for the individual settings. Additionally because many of the instruments could be easily modified to more directly document individual accomplishments, they were modified and therefore not useful to the core.
Although the core grew out of a grass roots need, many of the sites already had different evaluation processes in place. The CETPs, who were just starting when the Core instruments and procedures were available, were much more willing to be participatory than the CETPs who were further along when the Core instruments and procedures were available. On the other hand, without the experience of the more mature CETPs the instruments and procedures would have taken much longer to produce.

There were no consequences for not participating. Quality of each CETP was not judged on its participation in the Core and participation in the core was not linked to funding. Additionally CETPs whose individual evaluations showed more accomplishments were praised.

**RECOMMENDATIONS**

- Plan for the program evaluation before funding begins
- Require participation in the evaluation
- Consider partial participation—either only some sites participating or some sites collecting only some data
- Have project evaluators in the loop from the beginning
- Substantively involve PIs
- Have the program evaluation data play a critical role in project evaluation

**REFERENCES**


