C. PROJECT DESCRIPTION

Evaluation Systems and Systems Evaluation:
Building Capacity and Tools for Enhancing STEM Education Evaluation

This research is designed to enhance the capability of the field of evaluation to develop more effective evaluation systems and to conduct more rigorous systems evaluation. By evaluation system we mean the comprehensive and integrated set of capabilities, resources, activities and support mechanisms for conducting evaluation work. By systems evaluation, we mean the assessment of the functions, products, outcomes and impacts of a set of programs, activities or interventions. The intellectual merit of this research stems from its contribution to the improvement of our understanding of the role of systems in evaluation specifically through (1) development of a formal model of evaluation system needs that provides structure for evaluation capacity building and delivery; (2) development of one or more measures of evaluation systems capacity; and (3) conducting a pilot test of an evaluation capacity-building system and measure(s). The broader impacts of this research stem from the way it will: strengthen the infrastructure for research and evaluation through the development of new measures for local evaluation capacity building ; and, develop a new evaluation network that will serve as a model for multi-institutional multi-program evaluation partnerships.

The specific context for this effort is the evaluation of Science, Technology, Engineering and Mathematics (STEM) education programs, but its implications extend to the field of evaluation generally. The research will result in a set of specific products that will have immediate utility in STEM education and be adaptable to non-STEM contexts, including a measurement approach and instrument(s) that can be used to assess and monitor evaluation capacity and support in a STEM outreach and educational system.

This project is directed at how we might develop more effective evaluation systems both generally and for STEM education programs in particular. The landscape of contemporary evaluation in recent years is replete with evaluation systems development efforts, yet we know surprisingly little about how such endeavors might best be developed and themselves evaluated. This project uses structured participatory mixed methods systems approaches to: identify the needs that STEM educators and evaluators have for evaluation of their programs; develop a formal model of an evaluation system to address those needs; identify the major components of that system and how they might work together; and, operationalize one or more measures of evaluation capacity building that can be used to assess the development of that evaluation system. The model and measures will be implemented and tested in a pilot project in New York City that includes both formal and informal STEM education programs. The project will contribute to our fundamental knowledge regarding the nature and structure of evaluation systems, how they can be developed more effectively and how to evaluate their development. While this particular project involves STEM education and outreach specifically, the completed models and measures for evaluation systems will be useful in many substantive areas that extend far beyond the STEM context. Thus, this project generates new knowledge in the evaluation of STEM education specifically and explores new concepts in the evaluation field generally. In addition, the model and measures proposed here will enhance the instrumentation, networks, and partnerships developed to support both STEM evaluation and systems of evaluation.

The research project proposed here, is related to the idea of systems and systems thinking. It will contribute to our understanding of how evaluation systems that are regional and include university, school and community components might best be developed and integrated, and will provide specific measures and tools of relevance for monitoring evaluation capacity. It also contributes to the development of an evaluation system, in this case, one that spans the STEM endeavor, and pilot tests these approaches in a local STEM education context. Doing systems work in evaluation requires methods and tools that are appropriate to that work. This research describes the idea of systems methods and utilizes in a creative and novel way several recently developed methods to accomplish the systems work proposed herein.
This type of effort is critically important at this moment for the field of evaluation generally and for STEM evaluation in particular. It will help the field of evaluation address systems-level challenges. Despite the growing demands for such work, evaluation models and theory are largely stuck in the tradition of discrete one-off evaluation projects and have only recently begun to address either the need for integration across organizational levels and structures, or suitable evaluation methods and measures for doing so. This proposed research is designed to put the issue of systems in evaluation more squarely on the map and provides both a conceptual framework for thinking about evaluation systems, and a set of specific methods and tools that will enhance our ability to accomplish high-quality system evaluation work.

Cross-Cutting Themes: Systems and Systems Methods

There are several key themes that are foundational to this proposal and warrant some introductory explanation.

**Evaluation Systems and Systems Evaluation.** One of the most striking characteristics of the evolution of evaluation in the past twenty years is the increasing degree to which evaluation encompasses the idea of “systems.” At the outset, we need to acknowledge the multiple ways that the term “system” is being used throughout this proposal. Standard dictionaries offer no less that 12 distinct definitions of “system,” its principle one being “a regularly interacting or interdependent group of items forming a unified whole.” A more formal definition is that a system “may be defined as a set of elements standing in interrelation among themselves and with the environment” (Bertalanffy, 1995). These definitions mask a considerable definitional complexity and a history of ideas that spans centuries. One ought to undertake research on systems topics with a fair amount of caution and concern that these definitional issues may muddy the waters and confuse the issues. In a very real sense, the challenge is how to investigate system issues without becoming bogged down in myriad debates and recent histories about such definitional battles.

In such circumstances, it is desirable to take a pragmatic and more direct approach to defining the territory. In our view, “systems” refers to parts, the whole, and their interrelationships. Systems are nested within other systems – systems constitute the parts of still larger systems. In this sense, virtually all evaluation already can be considered “systems” based – the complex endeavor of any evaluation, even the most defined and circumscribed program evaluation – is an interaction of many parts (myriad stakeholders, participants, measures, and organizing structures such as research designs, analytic procedures, and environmental and policy contexts) into an emergent whole that we call an evaluation.

While this may be the case, we take the position that the idea of a “system” is a continuum ranging from simpler systems and components to more complex systems of systems. In this proposal, when we use the term “system” we are not referring to traditional unitary evaluation projects or programs, but to multiple evaluations considered in concert, to multiple programs or initiatives, or both.

We make an important distinction in this research between **evaluation systems** and **systems evaluation**.

- **An evaluation system** is an arrangement of various components – values and principles, methods, resources, and structures – designed to conduct and support an evaluation capability. Usually we construct an evaluation system as a way to support many separate evaluation projects, and often with the express purpose of being able to integrate or compare the results of these distinct evaluations.

- **A system evaluation** is an evaluation of a system and its processes and effects. The system might be a collection of programs that address a common conceptual theme (as in STEM evaluation), a series of discrete studies in an area of research (as in meta-analysis) or an organizational entity (such as a set of research centers or school districts).

It is not necessarily the case that the two terms need always be paired. It is possible to develop an evaluation system to conduct an evaluation of a single program. One might reasonably argue that the increasingly complex evaluation demands made by participatory approaches, multiplistic principles and
mixed methods almost necessarily leads each traditional “simple” evaluation to being definable as a “system”. It is also reasonable to think of evaluating a system without constructing an evaluation system, as with a discrete study of a single major indicator of system performance over time. However, it is almost certainly the case that these two endeavors tend to engender one another, and accomplishing them well is both a major contemporary challenge in evaluation and a central theme of this proposal.

Over the past two decades, the process of evaluation has become increasingly complex. For instance, consider how conducting a traditional one-off evaluation of a program has evolved. In contrast to several decades ago, high-quality program evaluation today relies on the integration of a complex array of perspectives and methods including, but not limited to: development and integration of program theory (Chen & Rossi, 1990) and program logic (Kellogg Foundation, 2001); use of a critical multiplism (Shadish, 1986) of samples, measures, program implementation and analyses; integration of qualitative and quantitative mixed methods (Greene & Caracelli, 1997); and the incorporation of a variety of stakeholders and participants (Fetterman, Kaftarian, & Wandersman, 1996; Macaulay, 1999; O'Fallon & Deary, 2002; Reason & Bradbury, 2001). These demands for complexity in evaluation simply were not present, or were present in more incipient ill-defined form as recently as 1980. Consequently, there is a growing need to address integration of the complex activities and methods of even the simplest evaluation, and for tools and methods that can be used to design and implement evaluation systems.

The growth in complexity of evaluation processes has been matched by a commensurate growth in the complexity of what we are asked to evaluate. The emphasis on program evaluation has been expanded with increased calls for evaluating multiple programs, initiatives, portfolios, or directly, systems. The need to understand what works generally (rather than just in specific contexts or settings) has led to the rise of meta-analysis (Chalmers & Haynes, 1994; Lipsey & Wilson, 2000) and the desire to summarize research at higher levels of aggregation than just the individual program. Consider also the increased reporting and accounting needs of all levels of government - especially at the federal level – with corollary increased demands for evidence and evaluation of broad programs and research initiatives (Brainard, 2002). Most of the pressures for accountability existed in at least incipient forms as early as 1980, but the ways we approach addressing these evaluation demands (e.g., through meta-analysis, research syntheses, performance management) has evolved significantly since then. All of these developments suggest that we are increasingly being called upon to conduct not just program evaluation, but more integrated system evaluation.

The need for systems thinking in evaluation and in the many fields within which it is used has increased markedly in recent years. It has been a major topic in the profession of evaluation, being addressed increasingly at national conferences and in the development of special interest groups such as the Systems in Evaluation TIG of the American Evaluation Association. Several books by prominent evaluators (Michael Patton, Bob Williams) on systems in evaluation are currently being developed. Systems efforts in substantive fields are well illustrated by looking at the field of public health. There, increasingly complex systems interventions have become normative in critical substantive areas like tobacco control (the COMMIT and ASSIST evaluations). Major reports (Institute of Medicine, 2001) emphasize the need for system approaches and methods, and explicitly describe the critical role that evaluation must play in such endeavors. Significant efforts at systems evaluation, especially through evidence-based practice, can be seen in the Community Guide (Briss et al., 2000; Pappoianou & Evans, 1998), Cancer Control PLANET, and the evolution of the Syndemics initiative at the Centers for Disease Control and Prevention and the Initiative for the Study and Implementation of Systems (ISIS) of the National Institutes of Health (in which the PI for this project has played a major role). A special issue of the American Journal of Public Health on systems thinking and modeling is scheduled to be released in Fall of 2005. This kind of evolution, emerging in recent years in many of the fields in which evaluation is practiced, suggests that the general area of systems evaluation is a timely and important research topic.

**Systems Methods and Approaches.** As we shift to more systems level endeavors in evaluation, there is a need for methods and tools that are explicitly designed to handle the complexity of systems work. By
**systems methods** we mean any methods that were designed or can be adapted to operate in systems contexts. Because of the nature of systems work, these methods often are participatory and collaborative. They might be extensions of traditional methods, as in the use of survey research or traditional measurement tools applied to systems contexts. But increasingly these methods are themselves more complex hybrids and mixtures of traditional methods, participatory processes and novel uses of technology. An example of such a methodology that will be used in this research is the method of structured conceptualization or collaborative concept mapping developed by the PI of this proposal (Trochim, 1989c; Trochim & Linton, 1986; Trochim, Milstein, Wood, Jackson, & Pressler, 2004). The method was explicitly designed to enhance the ability of a system to develop shared frameworks and models, and measures or observational approaches for assessment and evaluation. *A methodological agenda for this research is to advance our understanding and encourage development of appropriate systems methods and approaches, either through adapting traditional methods more effectively to systems contexts or by suggesting development of new methods.*

**Encouraging a Transformative Outreach System.** This research is set within a larger theoretical context of transformative outreach and education that we have been developing at Cornell that for STEM education would integrate educational, school-based and community practice components with university-based outreach and extension elements. The research proposed here will encourage development of this model by significantly advancing our understanding of the role of evaluation in transformative outreach. The central idea is that a university’s outreach efforts could have a transformative impact on STEM teaching and learning if critical dynamic components and interactions were added to create a system out of what are now typically a myriad collection of distinct and loosely coupled programs. Evaluation has a critical and largely underexplored role to play in developing such a theoretical framework. The idea of transformative outreach draws on several important existing research literatures, most notably those of youth development and of transformative learning, and advances several important principles:

- **Outreach should be guided by the needs of educators and schools, determined through a carefully designed interactive process and regularly updated.** It should operate in the domain where those needs converge with university resources. This research will contribute by utilizing structured systems methods to assess needs, match them with resources, encourage shared capacity building and assess progress.

- **No single university outreach program by itself can adequately support all needed elements.** Coordination among outreach programs is essential both for efficiency and effectiveness. In this research we will model how such collaboration can occur specifically for evaluation.

- **General university efforts to recruit and train underrepresented students, staff and faculty must be integrated with efforts of specific programs such as STEM.** Currently they generally are not at most universities. In this research we will explicitly encourage such integration in the specific context of STEM evaluation.

- **To have a transformational impact, resources must be focused, both geographically and in content and goals.** Offering an assortment of resources to various users is laudable but unlikely to make large and lasting improvements. Formal partnerships should be established that include schools, community organizations, university administrative units and educational and outreach programs to focus outreach with the aim of transforming the teaching and learning of STEM subjects in partner schools. In order to reach a more racially and ethnically diverse group of students, Cornell must reach beyond its local communities to one or more large cities. For this research, New York City provides a compelling context that enables us to involve both a school (formal) and a community-based (informal) local STEM program and integrate our efforts with the system of Cornell Cooperative Extension through the local office there.
To achieve a transformational impact, innovations from the latest research in systems thinking and modeling should be incorporated. Relevant innovations from systems thinking, including from the area of complex adaptive systems (CAS) research (J. Holland, 1995; J. H. Holland, 1998) include: processes that encourage interactivity and direct collaborative networking experiences between groups that do not typically interact (e.g., scientists and administrators with teachers and community educators); collaborative identification of a shared emergent model of the problem or issue at hand (e.g., modeling evaluation needs); the identification of simple evaluation procedures (“rules”) which, when followed by different stakeholders (“autonomous agents”), encourage greater diversity and adaptivity in the system; and the provision of a coherent and timely feedback mechanism to the system through monitoring and evaluation. This research incorporates all of these systems thinking approaches.

Major Project Activities

Systems that consist of multiple projects and organizations typically require a wide range of evaluation skills, resources and capabilities in order to provide the essential feedback mechanisms for system learning and to meet the accountability reporting needs of their funders and stakeholders. Some of these needs are best met by developing internal organizational or system capacity as by hiring experienced evaluators, supporting external or in-service training in evaluation, hiring external consultants, and so on. Evaluation capability is sometimes met by funders who require evaluation, through additional funding for this function or through evaluation technical assistance. In many cases individuals or groups within organizations simply seek out resources and information on evaluation, increasingly these days through technology-based solutions such as evaluation websites. There is no shortage of useful information available for meeting evaluation capacity needs, including websites and materials that present case studies, measures, examples, tutorials, and other resources, available from public sector government agencies such as NSF (Frechtling, 2002; Frechtling & Sharp, 1997) and the Centers for Disease Control (www.cdc.gov/eval/), individual researchers (www.socialresearchmethods.net), professional associations (www.eval.org) and foundations (www.wkkf.org). The problem isn’t a lack of resources; rather, there are too many resources of varying and indeterminate quality that are not screened or adapted for use in particular contexts. And, even if program staff can get to high quality resources, their ability to use them effectively is greatly enhanced by appropriate support systems and experienced evaluators. There is little in the way of systematizing the plethora of resources and little support for practitioners who need to navigate this complex terrain. Nor is this problem solved through developing and providing more resources, more websites, and more materials. Despite the ubiquitous nature of evaluation needs, and the abundance of supporting resources, there is surprisingly little attention paid in the evaluation literature to the development of systems for supporting evaluation, and especially a paucity of research on how to coordinate and integrate the varied existing supports of the evaluation function. Much of the evaluation literature in this area emphasizes developing the government’s evaluation capacity (Wye & Sonnichsen, 1992), rather than enhancing the practitioner systems (Compton, Baizerman, & Stockdill, 2002). This project is designed to: develop a model of STEM evaluation system needs and how they might be addressed; develop one or more measures of evaluation capacity that can be used to assess change efforts; and, pilot test a capacity-building effort in two local STEM programs while simultaneously testing the evaluation capacity building assessment methods developed.

The specific context for this project is the Cornell STEM Evaluation Network (see description of Project Partners below), an extensive collaboration of STEM education and evaluation partners that includes school, community, and university groups and organizations, integrated with a national network of STEM educators.

Activity 1: Develop a Model of STEM Evaluation System Needs and Capacity Building: The first year of this project will be devoted to developing a model of the needs of STEM educators for evaluation support and assistance and specific plans for how they might be addressed.
Participants. Partners in the Cornell STEM Evaluation Network will be engaged in this activity, including: the NRCEN STEM educators; Cornell CIBT and Garden Mosaic projects in New York City; STEM scientists, educators, evaluators and graduate students at Cornell from the NBTC, IGERT, CSIP, and EI programs; and, key faculty, staff and administrators of Cornell Cooperative Extension. For all web-based components of this activity, all identified persons in these groups will be invited to participate (with the exception of the unstructured sorting task described below, for which a subgroup will be selected). For the Model Development Workshop, representatives of each of these groups will be identified and invited to attend, with all others invited to participate through the web and/or videoconferencing.

Mapping STEM Evaluation Needs. Prior to the Model Building Conference, participants will conduct a web-based concept mapping project (Trochim, 1989a) to map comprehensively the needs for STEM evaluation support and capacity. Participants will begin by doing a web-based free listing or brainstorming (Coxon, 1999) task to identify specific needs for evaluation support in STEM contexts. It is expected that several hundred statements will be generated. These will be edited and synthesized using a type of content analysis (Krippendorf, 2004) to reduce the final statement set to between 80-100 statements. A sampled heterogeneous subgroup of participants will use the web to do a free sorting of these statements (Coxon, 1999) and all participants will rate them for relative importance for effective STEM evaluation support. In addition, participants with expertise in evaluation methods will be asked to rate each statement for relative feasibility from an evaluation perspective. The standard statistical analysis for concept mapping (Trochim, 1989a, 2004) will be applied to these data. This involves: the aggregation of the sorting data (Weller & Romney, 1988) into a matrix of similarities; analysis of the similarities using two-dimensional multidimensional scaling (Davison, 1983; Kruskal & Wish, 1978); and, analysis of the resulting coordinate matrix by hierarchical cluster analysis using Ward’s algorithm (Anderberg, 1973; Everitt, 1980). The MDS configuration of the statement points is graphed in two dimensions in a "point map" that displays the location of all the brainstormed statements with those closer to each other generally expected to be more similar in meaning. A "cluster map" is also produced that partitions the statement points based on cluster analysis results. Figure 1 shows an example of an integrated point and cluster map that constituted a model for planning and evaluating comprehensive public health services (Trochim et al., 2004). The points on the map are the ideas that were brainstormed through a web-based process, with locations determined through multidimensional scaling of participant sort data. Hierarchical cluster analysis partitioned the ideas into seven clusters that the participants named in an interpretation session. These were arranged by participants into a higher-level model of systems and community factors crossed with structure, infrastructure and transmission components.

![Figure 1. Example of an integrated point and cluster concept map model (Trochim et al., 2004).](image-url)
The 1-to-5 importance and feasibility rating data are averaged across persons for each statement and each cluster. This rating information is depicted graphically in a "point rating map" showing the original point map with the average rating per item displayed as vertical columns in the third dimension and, in a "cluster rating map" that shows the cluster average rating using the third dimension.

Pattern matching (Trochim, 1985, 1989b) will be used to explore consensus regarding the relative importance of the statements across different subgroups (e.g., STEM educators, teachers, school administrators, scientists, educators, evaluation methodologists) and to identify evaluation needs that are both perceived as important and judged as feasible to address. Bivariate plots of statements by importance and feasibility will enable identification of the evaluation needs that are perceived as both high in importance and most feasible to accomplish. Figure 2 shows an example pattern match (left) and bivariate analysis (right) for the public health project model in Figure 1. Appropriate statistical testing for reliability of data and for differences among groups will also be accomplished, including correlational analysis and Analysis of Variance (ANOVA) for pattern matching data.

Figure 2. Pattern match (left) and bivariate analysis (right) within cluster from public health model concept mapping (Trochim et al., 2004).

There are several distinct advantages to using concept mapping in concert with the model building conference. First, concept mapping enables broad-based national participation of STEM educators and others in the Cornell STEM Evaluation Network. Second, it imposes little burden on participants while yielding considerable detail. Participants take part asynchronously at their convenience; total time on the project over several months will be between 30 and 90 minutes over several sessions depending on the group the person is in. Third, concept mapping is an ideal methodology for needs assessment, especially with geographically disbursed populations (Witkin & Altschuld, 1995). Finally, concept mapping is a mature methodology and technology that has been used in a wide variety of similar model building contexts including: for developing conceptual models for complex constructs (Daughtry & Kunkel, 1993; DeRidder, Depla, Severens, & Malsch, 1997; Trochim, Stillman, Clark, & Schmitt, 2003; White & Farrell, 2001); for identifying perspectives of diverse stakeholders (Donnelly, Donnelly, & Grohman, 2000; Southern, Young, Dunt, Appleby, & Batterham, 2002; Trochim, Cook, & Setze, 1994); and, for planning and evaluation (Biegel, Johnsen, & Shafran, 1997; Johnsen, Biegel, & Shafran, 2000; Shern, Trochim, & Christina, 1995).

Model Building Conference. The web-based concept mapping project will be used as the starting point for a STEM evaluation model building conference designed to interpret the map results, develop a model of evaluation needs, and begin developing both a strategy for addressing needs through capacity building, and measures of evaluation capacity building for assessing such efforts. The purposes of the proposed conference are to:

- develop a model of the needs for evaluation of STEM outreach;
- identify current resources and best practices in the evaluation of STEM outreach that could be directed to addressing those needs;
• identify potential gaps in the existing resources and support system for evaluation;
• begin development of a measurement system for assessing STEM evaluation capacity and evaluation system performance (see Activity 2); and
• begin development of a system for building evaluation capacity for enhancing local STEM educational program evaluation capability (see Activity 3).

The participants in this conference will be a heterogeneous subsample of the partner groups for this project. We intend to invite ten participants from NRCEN to represent nationally distributed STEM programs, ten from STEM programs within Cornell, and five from New York City partners to form the core group of participants at the conference. Format for the two-day conference will be a combination of facilitated group activities, presentations, discussions and breakout groups. We will structure the meeting using a combination of videoconferencing and webcasting to enable participation of those who would like to take part but were not able to be invited to the conference. Interactive videoconferencing will link the Cornell campus in Ithaca with the Cornell Cooperative Extension office in New York City, enabling partners there to participate. We will also make the entire conference available in real time using webcasting. Such high intensity broadly-oriented technology-based and face-to-face interaction among a heterogeneous array of stakeholders in STEM education and its evaluation is consistent both with the goals of funders for such networking and with contemporary systems thinking that sees such efforts as essential for system capacity building. It will have the direct effect of accomplishing tasks relevant to this activity and the perhaps even more profound indirect system effect of engendering new network connections among the groups involved.

Activity 2: Develop Measures of Evaluation System Capacity. At the Model Building Conference an Evaluation Systems Capacity Working Group of 10-15 people will be formed from the participants and charged with the responsibility for guiding project staff in developing one or more draft measures as appropriate for assessing evaluation capacity in the STEM system. The evaluation needs model and conference discussion will be the starting point for this measurement development effort, and the group will work by teleconference subsequently to accomplish this work. The process for measurement development will include the following steps:

• For each cluster/topic in the model, the statements from the map will be used as catalysts for suggesting specific operationalizations for that topic. Working group members will brainstorm additional potential ways to measure each topic, assisted by staff evaluation methods experts.
• Potential measures/indicators will be classified by type of measurement (e.g., survey, scale, index, observation, records analysis, etc.).
• Each potential measure will be prioritized using information about the relative importance of the construct and feasibility of implementation.
• Draft measurement protocols (instruments, instructions for collecting data) will be produced by project staff and reviewed by the working group.

The measures will be pilot-tested in the initial stages of implementation of the Pilot Project (Activity 3).

Activity 3: Pilot Test Evaluation Capacity Building System and Measure(s). The purpose of this activity is to test the measure(s) developed in Activity 2 in a local program context where we intervene to improve evaluation capacity. There will be several steps in this process:

• The Evaluation Systems Capacity Building Working Group will use the model developed in Activity 1 and the information about available evaluation resources and gaps to advise project staff on development of a system for addressing evaluation needs in the two New York City STEM program sites.
• The measure(s) will be administered to obtain a baseline assessment of evaluation capacity. The performance of the measure(s) will be assessed, including checks on reliability, consistency and validity as appropriate for the type of measure(s) developed. The measure(s) will be revised based on this initial administration.

• The evaluation capacity building system will be implemented to the extent possible over a 6-9 month period in the two local programs.

• The revised measure(s) will be administered a second time, and appropriate checks on reliability, consistency and validity conducted. In addition, for elements of the measure(s) that remain unchanged, we will pilot the use of the measure(s) for assessing change in evaluation capacity.

The primary purpose of this activity is to pilot test both the capacity building system and the measure(s). It is not intended to be an evaluation of the capacity building effort per se, although it will likely provide useful feedback. Instead, the pilot test is being undertaken to help assure that the intervention and its measure(s) will be ready for wider scale implementation subsequent to this research.

**Project Partners: The Cornell STEM Evaluation Network**

A network of partner organizations and programs has been organized specifically to accomplish this research. The development of such a network is consistent with funder’s goals and is itself a broader impact of this research. Letters of support from participating individuals and organizations accompany this proposal.

At the national level in STEM education, this project will involve the NSF Research Center Educators Network (NRCEN) (www.nrcen.org) which was established in 2001 to bring together educators from around the country in Materials Research Science and Engineering Centers, Engineering Research Centers, and Science and Technology Centers. The inaugural meeting of this group was held at Cornell University, and a vibrant network of educators that now includes over one hundred participants grew out of this effort. An annual meeting has been held each year since 2001, with a focus on connecting centers together to have a broader impact on STEM education nationally through collaboration. This group of educators will provide a national perspective and deep local experience on evaluation issues facing the STEM education community.

At the school and community levels, we will concentrate our involvement in this project with the New York City sites of two specific Cornell-based STEM programs, one a formal school-based program and the other an informal community-based one:

• The Cornell Institute for Biology Teachers (CIBT) (http://cibt.bio.cornell.edu/) includes elementary through high school teachers and school administrators. CIBT develops classroom resources and activities and offers unique opportunities for professional development. We will work with the extensive CIBT project in their New York City site.

• The Cornell Garden Mosaics project (http://www.gardenmosaics.cornell.edu/) is an NSF-funded science education program that combines intergenerational learning, community action and learning about different cultures. Youth learn from elder gardeners who share their planting practices and cultural backgrounds, and from science resources developed and made available by the program. We will collaborate with the New York City site.

These programs will constitute the pilot sites for the evaluation capacity-building and assessment effort of this project.

For the university context, we will concentrate our efforts at Cornell University, an ideal real-world laboratory for investigating systems approaches for STEM contexts. Cornell is one of the premier research institutions and has one of the largest systems of creative and productive education programs aimed at
improving K-12 education in science, technology, engineering, and math (STEM). The range and depth of programs and outreach activities may be sampled by visiting the “Learning, Teaching, and Continuing Education” page on Cornell’s Outreach web site (http://www.cornell.edu/outreach/programs.cfm#learning). The co-PI for this proposal (Hamilton) is the Associate Provost for Outreach at Cornell, with primary responsibilities for this system.

We will also involve in this research several of the major existing STEM groups at Cornell – the following, all of them funded by NSF, have agreed to participate:

- The NSF-funded Cornell Nanobiotechnology Center (NBTC) (http://www.nbtc.cornell.edu/) is a highly interdisciplinary endeavor that features a close collaboration between life scientists, physical scientists, and engineers. It has a fully integrated education and outreach effort in which all NBTC faculty participate. The Center brings together experts in their fields from Cornell University, the Wadsworth Center (New York State Health Department in Albany), Princeton University, Oregon Health & Science University, Clark Atlanta University, and Howard University. It also involves the active collaboration of K-12 educators, the Sciencenter Museum in Ithaca, NY, and representatives from industry and the government.

- The NSF-funded Cornell Integrative Graduate Education and Research Traineeship (IGERT) program (http://www.chaos.cornell.edu/) trains PhD scientists and engineers with the interdisciplinary background and the technical, professional and personal skills needed to address the global questions of the future. Through the use of innovative curricula and internships, and by focusing on problem-centered training, these programs give their graduates the edge needed to become leaders in their chosen fields. We will involve students and faculty with responsibilities for educational efforts through IGERT.

- The Cornell Science Inquiry Partnerships (CSIP) is an NSF Graduate Teaching Fellows in K-12 Education project (http://csip.cornell.edu/) that provides fellowships to Cornell graduate students who spend an average of 15 hours per week teaching collaboratively with secondary teachers in about 40 rural and urban schools within 100 miles of Cornell. CSIP has funded 50 graduate student fellows over the past five years.

- Environmental Inquiry (EI) is an NSF Instructional Materials Development project (http://ei.cornell.edu/) through which Cornell scientists work with secondary teachers to develop curriculum materials designed to enable secondary students to conduct original experiments using research protocols similar to those used by professional scientists. This work led to a series of four books published by the National Science Teachers Association.

These partners will enable us to integrate the views and ideas of STEM scientists, educators and graduate students. Through their involvement in this research we will be encouraging and training a cohort of graduate students in evaluation issues generally and STEM evaluation in particular.

Finally, the fact that Cornell is a Land Grant Institution provides a unique opportunity to integrate the system of Cornell Cooperative Extension (CCE) (http://www.cce.cornell.edu/) in this research and in STEM education evaluation generally. The Cooperative Extension system is historically one of the most effective educational outreach efforts (Peters, Jordan, Adamek, & Alter, 2005; Peters, Jordan, Alter, & Bridger, 2003), but has been a largely untapped resource for STEM educational evaluation, despite their integral involvement in STEM education efforts such as the Garden Mosaic program that will be a partner for this research. The Cornell Cooperative Extension educational system enables people to improve their lives and communities through partnerships that put experience and research knowledge to work in a variety of substantive areas. CCE has over 400 extension educators located in every county in New York State; 245 faculty and staff in Cornell's New York State Colleges of Agriculture and Life Sciences, Human Ecology, and Veterinary Medicine; and over 50,000 volunteers participating in both program and organizational leadership. Cornell is considered a premier institution among the 103 land-grant colleges.
and universities across the United States. The PI for this project (Trochim) has major responsibilities for CCE evaluation. The Director of the CCE Office in New York City has experience in evaluation and will play a major role in facilitating the coordination of the Cornell campus efforts and the project in New York City. The leadership of CCE and the Deans of the relevant Colleges at Cornell see this project as one that can be a model for integrating Extension nationally into STEM education and can have a broader impact by enhancing evaluation capabilities in the extension system generally.

Focusing on a single university context and its collaborators is especially appropriate for this research. Universities are central nodes of the national system of scientific research and play a critical role in STEM education. The specific context we are focusing on at Cornell University is integrally linked to national efforts both through participation of the network of STEM educators (NRCEN), and through involvement of scientists and educators at Cornell sites who participate in national research and education initiatives funded by NSF and others. Studying this system will enable development of models and tools that are directly adaptable by and have broader impact for other local, regional and university-based efforts in STEM education. At the same time, focusing the effort on Cornell and its linked collaborators enables this research to be more cost effective and feasible within the timeframe of this proposal.

Consistent with the priorities of the funders and the commitments of those responsible for outreach, all of the STEM programs and entities engaged in this project share two primary principles. The first is an emphasis on inquiry as an approach to teaching and learning STEM, in contrast to more conventional didactic approaches. Many of the programs devise hands-on activities tied to curricular goals, demonstrate them in classrooms, teach teachers how to use such activities and invent or adapt them, and provide kits of materials to enable teachers to conduct these activities in their classrooms. The second theme is equity, with a concerted focus on stimulating interest and enhancing performance in STEM-related subjects among girls, low-income students, and students from racial and ethnic groups who are under-represented in STEM fields, and in encouraging the training of under-represented groups in both education and evaluation skills.

Project Staff

**William Trochim (PI),** is a Professor in the Department of Policy Analysis and Management at Cornell University and is Director of Evaluation for Extension and Outreach. He received his PhD from the Department of Psychology at Northwestern University in the area of Methodology and Evaluation Research. His research is broadly in the area of applied social research methodology, with an emphasis on program planning and evaluation methods. Among experimentalists, he is known for his work in quasi-experimental alternatives to randomized experimental designs, especially the regression discontinuity and regression point displacement designs. In terms of research theory, he has extended the theory of validity through his articulation and investigation of the idea of pattern matching. In multivariate and applied contexts he is recognized for the development of a multivariate form of structured conceptual mapping, a general method for mapping the ideas of a group of people on any topic of interest that integrates traditional group processes (e.g., brainstorming, Delphi, focus groups, nominal group technique) with multivariate statistical methods (e.g., multidimensional scaling, hierarchical cluster analysis). He has written several books, including a widely used introductory research methods text, and articles that have appeared in the American Journal of Evaluation, New Directions for Program Evaluation, Evaluation and Program Planning, Evaluation Review, Journal of Clinical Epidemiology, Consulting and Clinical Psychology, Controlled Clinical Trials, Performance Improvement, and Medical Decision Making, among others. He is the developer of the concept mapping methodology and software that is offered through Concept Systems Incorporated. He has also been an active member of the American Evaluation Association, serving multiple terms on its Board.

**Stephen F. Hamilton (co-PI),** is the Associate Provost for Outreach at Cornell University, whose primary responsibilities are to facilitate exchange and collaboration among faculty committed to outreach, extension and public service; expand and extend new models for outreach; increase university wide
recognition of outreach; promote and facilitate evaluation of outreach and encourage the highest standards in outreach and extension; facilitate access to Cornell’s resources in conjunction with other university offices; and develop new mechanisms for communications between external stakeholders and the campus. His first priority in this role was to work on advancing the role and commitment of Cornell to K-12 education, particularly in science, technology, engineering and math. Hamilton is also Professor of Human Development and was named co-director of the Family Life Development Center in 1999. He has also served as department chair and chair of the Faculty Committee on Program Review. Dr. Hamilton’s research investigates adolescent development and education, with an emphasis on the interactive influences of school, community and work on the transition to adulthood.

**Plan of Work**

This research will take place over a two-year period. Figure 3 provides an overview of the work plan showing the major activities and subactivities of the project. Lines give expected duration of each activity. Circles represent specific products such as instruments or reports. All products will be made available on the project website.

![Figure 3. Two-Year Project Workplan Schedule.](image)

**Dissemination Plans**

Results from this research will be disseminated widely throughout the duration of the project. At the outset, a website will be developed to allow for distribution of up to date project materials, ongoing communication among the partners in the Cornell STEM Evaluation Network, to enable the concept mapping process, and to share the resulting model of evaluation needs of the STEM community. The measures developed will be made available to all participants in the project processes, and to the wider STEM and evaluation communities. A variety of reports and other projects will also be made available including: the Model Building Conference Report and the Project Final Report.
References


O'Fallon, L. R., & Dearry, A. (2002). Community-based participatory research as a tool to advance environmental health sciences. Environmental Health Perspectives, 110, 155-159.


