A Phase II trial of the Systems Evaluation Protocol for Assessing and Improving STEM Education Evaluation

Project Description

A. Introduction

“…despite decades of significant federal investment in science and math education, there is a general dearth of evidence of effective practices and activities in STEM education.” (p. 3)

Academic Competitiveness Council
(U.S. Department of Education, 2007)

Science, Technology, Engineering and Mathematics (STEM) education is at a critical juncture. The conclusion cited above results from a major federal interagency report headed by the U.S. Secretary of Education and completed in the past year. The challenge that is raised is essentially an evaluative one: How can STEM education programs provide evidence of the effectiveness of their practices and activities? The central purpose of this large empirical research project is to help address this evaluative challenge through the rigorous multi-site testing of a standardized STEM education evaluation protocol that has the potential to transform how evaluation in this field is conceptualized, accomplished and synthesized.

Several major forces are converging upon and posing a unique challenge to the STEM education endeavor.

The educational environment has changed significantly. The field of education has been significantly influenced by federal education requirements engendered by the No Child Left Behind (NCLB) Act of 2001 (Public Law 107-110) that imposed stricter teacher qualification requirements, led to increased emphasis on student standardized achievement testing, and called for schools to use “scientifically based research” strategies as the foundation for their educational programming. Regulations proposed and enacted in the U.S. Department of Education (U.S. Department of Education, 2003, 2005) have been especially controversial and have important implications for STEM education evaluation (Julnes & Rog, 2007).

There is increased demand for evaluation of federal programs. There is a greatly increased emphasis on the accountability of federal programs generally and on the need for comprehensive evaluation of such programs, including large scientific research grants and their education and outreach efforts (Brainard, 2002a, 2002b; U.S. General Accounting Office, 2000). For example, the Office of Management and Budget instituted the Program Assessment Rating Tool (PART) as part of the 1993 Government Performance and Results Act (GPRA) (Office of Management and Budget (OMB), 1993) and now requires every federally funded program to be reviewed on a regular basis, including an assessment of the quality of their evaluation and of the program’s functioning and effectiveness. As recently as November 2007, the U.S. Office of the President issued an Executive Order on Improving Government Program Performance that requires that all federal agencies establish for all federal programs that the “means for measurement of progress toward achievement of the goals are sufficiently rigorous and accurate” and that they should “assess performance of each program administered in whole or in part by the agency” (U.S. Office of the President, 2007).

There is increased demand for evaluation of STEM education programs. These changes in the federal environment raise profound challenges for the field of evaluation generally and for STEM education evaluation in particular. The Deficit Reduction Act of 2005 (P.L. 109-171) established the Academic Competitiveness Council in which officials from several federal agencies with STEM education programs over the course of a year created an overall inventory of 105 federal STEM education programs that constituted approximately $3.12 billion in total funding for Fiscal Year (FY) 2006. The ACC’s report (U.S. Department of Education, 2007) concluded that, “despite decades of significant
federal investment in science and math education, there is a general dearth of evidence of effective practices and activities in STEM education.” (p. 3) and their report recommended that: an inventory of goals and metrics for STEM education programs be developed and maintained; federal agencies should foster “improved evaluation and-or implementation of proven-effective, research based instructional materials and methods,” “improve the coordination of their K–12 STEM education programs with states and local school systems,” and “adjust program designs and operations so that programs can be assessed and measurable results can be achieved”; and, federal funding for STEM education “should not increase unless a plan for rigorous, independent evaluation is in place.” (U.S. Department of Education, 2007)

These forces potentially threaten STEM education as it is currently pursued, while at the same time offering the possibility for transforming it significantly in the 21st century. This research is designed to anticipate these challenges by enhancing our ability to conduct high-quality rigorous evaluations of STEM education projects and programs.

B. Background Research

Research Foundations of the Systems Evaluation Protocol (SEP)

This proposal builds upon research being conducted under a current NSF grant (National Science Foundation. Evaluation Systems and Systems Evaluation: Building Capacity and Tools for Enhancing STEM Education Evaluation. HER/REC/EREC. NSF Grant #0535492. William Trochim, Principal Investigator. 10/1/2005 – 8/31/2008). The purpose of the current grant is to address the general problem of evaluation systems and systems evaluation in a STEM context. An evaluation system is an arrangement of the various components – the values and principles, methods, resources, and structures – needed to accomplish multiple evaluations. Usually we construct an evaluation system as a way to support many separate evaluation projects, often with the express purpose of being able to integrate and disseminate the results of these distinct evaluations. A systems evaluation is evaluation that is undertaken from a systems perspective, usually incorporating insights and approaches from systems thinking and systems theory. We argue that it is essential in contemporary evaluation that we develop evaluation systems that are informed by systems thinking and that we integrate this approach into the individual evaluations we conduct.

A major product of the research currently being conducted is what is termed here a Systems Evaluation Protocol (SEP). In this section we briefly review the SEP and its research foundations and discuss the current initial pilot study of its implementation. This proposal requests funding for the next phase of empirical research on the SEP in STEM contexts.

One of the foundational rubrics for the SEP comes from the fields of evolutionary theory and natural selection (Darwin, 1859; Mayr, 2001) and especially from evolutionary epistemology (Bradie & Harms, 2006; D. T. Campbell, 1974, 1988; Cziko & Campbell, 1990; Popper, 1973, 1985). The central thrust of this line of research is that our knowledge, including our macro-level knowledge of STEM education interventions and programs, evolves according to the principles of natural selection, the trial-and-error cycle of (blind) variation and selective retention. Over time, program variations are tried and survive or not according to current socially negotiated selection mechanisms. In the modern era, evaluation has increasingly assumed an important role in both the generation of project variations through formative approaches like conceptualization and program theory methods (Caracelli, 1989; Chen & Rossi, 1990; Kane & Trochim, 2006; Linstone & Turoff, 1975; Trochim, 1989; Trochim & Kane, 2005; Trochim & Linton, 1986) and in the retention of variants that perform well through summative approaches like experiments or outcome assessments (D. T. Campbell, 1969; D.T. Campbell & Stanley, 1963; T. D. Cook & Campbell, 1979; Shadish, Cook, & Campbell, 2002). This macro-level evolutionary theory (phylogeny) is accompanied by project specific ontogeny: like organisms, each project has its own lifecycle and tends to proceed through different phases or stages.

We can see these evolutionary impulses manifested in the model from clinical medicine of phased clinical trials (Chow & Liu, 2004; Pocock, 2004). This model assumes that new interventions progress through
different lifecycle phases (ontogeny) of maturation and testing. Typically, Phase I clinical trials are small sample exploratory studies that look at dosage levels, implementation difficulties and potential side-effects of the treatment. Phase II trials concentrate on whether the treatment yields measurable changes and examine relationships among observed outcomes. It is in the late Phase II or Phase III trials that more formal controlled designs like the randomized experiment are employed. Phase IV trials are concerned with fidelity of transfer and implementation of treatments in uncontrolled real-world contexts. At a macro level (phylogeny) it is assumed following natural selection models of evolution that most interventions do not survive through this lifecycle. In fact, nearly three-fourths of all treatments are abandoned before a Phase III randomized experiment is ever mounted (Mayo Clinic, 2007).

We also see macro-level evolutionary thinking in the literatures, many of which originated in medicine, on meta-analysis (Hasselblad, 1998; Koretz, 2002; Smith & Glass, 1977; Sohn, 1996; Wolf, 1986), research syntheses and systematic reviews (Chalmers & Haynes, 1994; D. J. Cook, Mulrow, & Haynes, 1997; Davies, 2000; Dickersin, 2002; Dickersin, Scherer, & Lefebvre, 1994; Evans & Benefield, 2001; Gallagher, 1999; Juni, Altman, & Egger, 2001; Silagy, Middleton, & Hopewell, 2002; Silagy, Stead, & Lancaster, 2001; Wortman, Smyth, Langenbrunner, & Yeaton, 1998) and evidence-based practice (Antes, 1998; Artelt, 2001; Brownson, Baker, Leet, & Gillespie, 2002; Cronje & Fullan, 2003; Davies, 1999; Dopson, Locock, Gabbay, Ferlie, & Fitzgerald, 2003; Eitel & Steiner, 1999; Gibbs, 2003; McAlister, Graham, Karr, & Laupacis, 1999; Montori & Guyatt, 2002; Pirrie, 2001; Procopis, 2002; Sackett, Richardson, Rosenberg, & Haynes, 1997). In all of these, the assumption is that most programs (organisms) will not survive over the long run and that the trial-and-error process will yield a residual evidence base of those that survive the increasingly stringent natural selection mechanisms of evaluation and experimentation.

This fundamental evolutionary notion and its manifestation in clinical medicine has become a powerful model in contemporary science and has influenced evaluation, including STEM evaluation. For instance, we can already see its central tenets in the recent ACC report on STEM education:

“There is a critical pathway for the development of successful educational interventions and activities, starting generally with small-scale studies to test new ideas and generate hypotheses, leading to increasingly larger and more rigorous studies to test the effect of a given intervention or activity on a variety of students and in a variety of settings. Different research methodologies are used along the development pathway, and corresponding evaluation strategies must be used to assess their progress.” (p. 13) (U.S. Department of Education, 2007)

These notions are critical to the SEP and to the purpose of this study. The SEP was explicitly designed to place projects on a lifecycle, to encourage a progression of educational projects through their lifecycles and to assure that evaluation approaches are appropriate to each phase. In this proposal, the SEP itself is the “project” or intervention. As with any project, it is subject to the same process of natural selection and we use this study as an opportunity to move the study of the SEP to a second phase of testing and at the same time to model the process of phased trials that is advocated in this approach.

The Systems Evaluation Protocol (SEP)

The idea of an evaluation protocol comes from research in clinical medicine. A protocol is the foundation of any clinical trial. It describes the standardized series of steps taken to accomplish the trial:

“Every clinical trial must have a protocol, or action plan that describes what will be done in the study, how it will be conducted, and why each part of the study is necessary - including details such as the criteria for patient participation, the schedule of tests, procedures, and medications, and the length of the study” (National Institutes of Health, 2007).

The SEP is a protocol that is designed to generate evaluation protocols or plans. It is a series of repeatable steps that when followed lead to the creation of project logic and pathway models and an evaluation plan that can subsequently be implemented and utilized.

Evaluation Foundations. The SEP was explicitly constructed to incorporate key elements from the full spectrum of contemporary evaluation theory including but not limited to: development and integration of
program theory (Chen & Rossi, 1990) and program logic (Kellogg Foundation, 2001); integration of
qualitative and quantitative mixed methods (Greene & Caracelli, 1997); the incorporation of a variety
of stakeholders and participants (Fetterman, Kaftarian, & Wandersman, 1996; Macaulay, 1999; O’Fallon &
Dearry, 2002; Reason & Bradbury, 2001); formative and summative designs (Scriven, 1967); non-
experimental, quasi-experimental and randomized experimental designs (D.T. Campbell & Stanley, 1963;
T. D. Cook & Campbell, 1979; Shadish et al., 2002); and so on. It is consistent with existing widely used
evaluation models (Centers for Disease Control and Prevention, 1999) and with approaches currently used
in STEM evaluation (Frechtling, 2002). Paradoxically, the SEP is a standardized protocol that enables
any project to develop a unique tailored evaluation plan. In this sense it addresses well the STEM
education environment which needs standardization of evaluation approaches while recognizing the
enormous varieties of contexts within which STEM programming occurs.

Systems Theory Foundations. While the SEP yields evaluation plans that meet current best practices, it
was also explicitly constructed to extend current evaluation practice through deliberate integration of
principles from systems theory. The topic of systems thinking and systems theory is extremely complex;
the literatures that informed the formulation of the SEP were considerable and included: causal feedback
(Richardson, 1991); stock–flow structures and open and closed systems (J. D. Sterman, 2000);
centralized, decentralized, heterarchical, hierarchical, and self-organizing systems (Kauffman, 1993,
1995); nonlinear systems and chaos (Strogatz, 1994); complex adaptive systems (Gell-Mann, 1995, 2003;
J. Holland, 1995; Waldrop, 1992); boundary conditions and scaling, (Strogatz, 1994; Yeomans, 1992);
emergence (J. H. Holland, 1998; Johnson, 2001); cellular automata (Wolfram, 2002); fractal self-
similarity (McKelvey, 1999); general systems theory (Bertalanffy, 1951); cybernetics (Francois, 2004;
Young, 1969); control theory (Sontag, 1998); information theory (Shannon, 1948); computational
simulation (Gilbert & Troitzsch, 1999; Resnick, 1994); system dynamics (Forrester, 1961, 1994, 1997;
Richardson, 1996; Society, 2002; J. Sterman, 2001); evolutionary theory, biology, and ecology (Capra,
1997, 2002; Capra, Mill Valley School District (Mill Valley Calif.), & Elmwood Institute., 1994; Capra,
Steindl-Rast, & New Dimensions Foundation, 1993); small world phenomena (Strogatz, 2003; Watts,
1999a, 1999b); and set, graph, and network theory (Capra, 2002; Capra et al., 1994; Maturana & Varela,
1980; Newman, 2003; Strogatz, 2003; Watts, 1999a, 2003). Because the SEP integrated principles
associated with these theories it helps to assure that projects that use it will incorporate such principles
when developing project pathway models and identifying key pathways and nodes (outputs and
outcomes); determining the boundary conditions for their project model, assessing project lifecycles, and
selecting evaluation designs that are appropriate to their project evolution.

Cyberinfrastructure Foundations. The SEP also builds on principles from contemporary
cyberinfrastructure research (National Science Foundation, 2007). While the SEP can be implemented as
a manual process that does not depend on any specific technology platform, it is designed so that it can be
enhanced throughout by using a system developed in our current research called the Netway, a Web 2.0
application consistent with second-generation web-based communities and hosted services such as social-
 networking sites, wikis, and blogs (Wikipedia, 2007). The Netway is constructed so that when project
educators and managers enter project information about activities, outputs and outcomes, the system can
immediately identify and suggest other existing projects that have similar or common elements and enable
the users to adopt or adapt these elements for their own projects while automatically creating networked
linkages of their models with others (Asim, Essegaier, & Kohli, 2000; Burke, 2000). Each new project
model adds to the online network of such models and can be accessed in turn by others. This also helps
ensure that different parts of the system can learn from each other and that even projects with no direct
contact with one another, can use the cyberinfrastructure to benefit from each other’s experiences
(Marathe, 1999). Evaluators who are supporting projects that use the Netway can see in real-time their
portfolio of project participants and what they are adding to the system, and can communicate with
project users about their models and evaluation plans, thus enabling new models of virtual consultation.
The system is designed so that researchers will be able to identify clusters of projects and programs that
are in their substantive areas of interests and learn about new and emerging programs and projects that are
responding to local needs and conditions and funders (e.g., NSF) can view meta-summaries of projects across program areas, see where they are in their developmental lifecycles, and more effectively manage their portfolios of evaluations. The Netway cyberinfrastructure is a creative incorporation of technology that fundamentally changes the nature of evaluation practice for both the evaluator and the practitioner. It has the potential to be a transformative mechanism for STEM evaluation particularly and for evaluation generally in the 21st century.

Phase I Pilot Testing of the SEP

Phase I exploratory pilot studies of the implementation of the SEP have been underway over the past two years with over fifty programs in nine organizations including two NSF-funded STEM sites (Cornell Center for Materials Research and Museum of the Earth) and seven Cornell Cooperative Extension offices throughout New York State (many of which have STEM education programs). The development, implementation, and testing of this approach included: forming partnerships between the systems evaluation team at Cornell and the partnering sites (Preparation), implementing the systems evaluation process (Logic Modeling), developing an evaluation plan (Evaluation Plan), providing a common web-based system for evaluation planning (Netway), and utilizing measurement tools to assess the systems evaluation process (Measurement Development).

Elements of the SEP. The SEP includes three broad stages, each of which involves several prescribed steps. The broad stages are Preparation, Logic Model, and Evaluation Plan, which once completed result in an evaluation plan that will guide the evaluation implementation. The complete draft SEP protocol can be found on the project website at http://evaluation.cce.cornell.edu/sep/DraftSEPProtocol.doc. Figure 1 provides a pictorial representation of the overall stages and accompanying steps.

![Figure 1. Overview of the Systems Evaluation Protocol (SEP) for planning and evaluation.](image)

Preparation. The planning stage of the SEP begins with a “Preparation” stage that is intended to acquaint the participants with the systems evaluation process and establish baseline information regarding the target program(s)/organization. To date the SEP has been implemented as a facilitated process including evaluation experts and partnering organizations. When an evaluation expert partners with an organization, they together form an Evaluation Planning Partnership (EPP).

The Preparation stage involves the following steps:
1. **Negotiate the Memorandum of Understanding**: The Memorandum of Understanding (MOU) describes the roles and responsibilities of participants in the Evaluation Planning Partnership (EPP). The document details the expectations for the evaluation consulting team, pilot site administrators and staff, and provides a timeline of project activities and completion.

2. **Assess Organizational Capacity**: The purpose of this step is to establish baseline, to get a sense of what is actually involved in running each of the projects, to determine the resources available within the organization, the degree of evaluation training the staff has already received, information technology resources, and organizational evaluation policies.

The preparation stage results in several deliverables including a contact list, Memorandum of Understanding, Program Checklist, and Organizational Evaluation Capacity Checklist.

**Logic Model.** The second stage, “Logic Model” is intended to further acquaint the participants with the SEP, enhance knowledge of evaluation concepts, use systems approaches to analyze the Evaluand, and begin developing program logic and pathway models.

The Logic Model stage involves the following steps:

1. **Stakeholder Analysis**: Determine all of the potential people and/or organizations that may have a stake in the evaluation.

2. **Introduction to Evaluation**: Introduce core evaluation concepts that will be needed to complete the SEP.

3. **Program Review**: Gain a firm understanding of the components and characteristics of the program including how it operates and whom it serves.

4. **Program Boundary Analysis**: Determine the conceptual limits of the program; what is “in” and what is “out” when defining the program.

5. **Lifecycle Analysis**: Determine the maturity of the program and how its level of evolution influences evaluation capacity and method choices.

6. **Logic Model Draft**: Generate an initial logic model including the assumptions, context, inputs, activities, outputs, short-, medium-, and long-term outcomes.

7. **Pathway Model**: Use the logic model as a basis for articulating clear and direct linkages between program activities and outcomes.

8. **Program-System Links**: Determine and analyze the causal links between the activities and outcomes and identify the “hubs” and key markers in the pathway model.

9. **Determining Evaluation Scope**: Determine the specific components of the pathway model that will be the focus in the upcoming evaluation cycle.

10. **Program Logic Model Synthesis**: Finalize the logic and pathway models including reviewing the program logic model, assessing the model from the perspectives of key stakeholders, reviewing the Program Boundary Analysis, reviewing the Program and Evaluation Lifecycle Analyses, and revising the models as needed. This step also involves turning integrating relevant research literature as it relates to the causal pathways that have been articulated in the Pathway Model.

The Logic Model stage results in several deliverables including a Map of Stakeholders, Lifecycle Charts, a Program Logic Model, a Program Pathway Model, and a Collection of Evidence.

**Evaluation Plan.** The third stage, “Evaluation Plan” involves several steps that aid in the creation of an evaluation plan that will serve as the protocol for the implementation of the evaluation. The Evaluation Plan stage involves the following steps:

1. **Introduce the Evaluation Plan**: Present and discuss the components of an evaluation plan.

2. **Develop Evaluation Questions**: Develop evaluation questions that are based on the logic and pathway models and will function as the driving questions at the core of the evaluation.
3. **Develop Sampling Plan:** Define the population of interest, sampling frame and sample and describe the source(s) of the evaluation data.

4. **Identify or Develop Measures:** Identify any measures already being used in evaluating the program and assess them for quality and feasibility; identify other potential existing measures that might be used; and, develop any new measures that are needed.

5. **Develop Evaluation Design:** Describe how the samples, interventions, and measures will be coordinated over time.

6. **Develop Analysis Plan:** Articulate the plan for analyzing the evaluation data.

7. **Develop Evaluation Reporting Plan:** Detail the plan for reporting the results of the evaluation to the key stakeholders identified earlier.

8. **Evaluation Schedule and Implementation Planning:** Develop the schedule for the evaluation and key implementation milestones.

9. **Finalize Evaluation Plan:** Review, finalize, and prepare to share the Evaluation Plan with leaders in the organization and other relevant stakeholders.

The Evaluation Plan stage results in several deliverables including a list of Evaluation Questions, a Collection of Measures, and a Systems Evaluation Plan.

**Netway: The Evaluation Cyberinfrastructure.** The Netway is a web-based cyberinfrastructure that was developed for and is being pilot tested in current NSF research. It constitutes a central common platform that can be utilized in the planning and management of evaluations across the entire spectrum of STEM education programs and is a central tool in the implementation of the SEP (Marathe, 1999). The Netway enables practitioners to enter information about an educational program (inputs, assumptions, contextual issues, activities, outputs and outcomes) and its evaluation (questions, participants, measures, design, analysis, reporting) to create a logic model, pathway model, and an evaluation plan. The term “Netway” is derived from the phrase “networked pathway” and refers to the system of programs or projects that is common in educational initiatives, and particularly in STEM initiatives at NSF. The system assumes that any program or project can be described with its own pathway model (a causal logic model). But from a systems perspective, these models are likely to be relatable – they are likely to share common activities, outputs or outcomes. When participants use the Netway to accomplish evaluation planning, the software enables them to identify these common components and thereby specify the system-level relationships among various and varied programs or projects. We term such a system a “networked pathway” model or “netway” model for short. These networks are created by connecting existing logic model information for a program and outcome measures to comparable information in other programs at other organizations in the Netway database.

**Measurement Development.** Several measures were developed in prior research to assess evaluation capacity in the STEM system: 1) An organizational capacity checklist to assess evaluation capacity before and after implementation of the SEP; 2) A program evaluation capacity checklist to assess specific programmatic characteristics; 3) Rubrics for assessing the quality of logic models and evaluation plans; and 4) Netway usage statistics. All of these have been pilot tested under current research and are being revised for use in this proposed work.

The previous NSF-funded research described here is still underway (scheduled completion date is August 2008). Findings to date have been presented at several national evaluation conferences and research publications are under development.

**C. Project Goals**

The proposed project constitutes one phase of a four-phase model for testing interventions and programs. This study is a Phase II trial. As such it is not meant to be a causal assessment of the effects of SEP; it is meant to assess whether the SEP program when implemented in several different common STEM contexts is associated with change in sort and intermediate outcomes related to evaluation and the quality
of evaluation. It is assumed that if the results of this trial are positive subsequent funding will be sought for a national Phase III effectiveness trial. This proposed trial will take place over a period of five years and will consist of several subprojects conducted by different collaborative multidisciplinary teams. This length of time is required because the intervention itself requires preparation time, several years to accomplish and additional time for short and intermediate effects to be detected and results to be analyzed.

Previous work by the project team constituted a Phase I study to develop and implement the initial version of the SEP model. The results from that study provided the foundation for this project. This project is intended to apply the results from prior research to the next phase of the model.

Specific goals for this project are:

1. Assess the degree to which the SEP is associated with changes in organizational evaluation capacity and performance outcomes.
2. Assess the performance of alternative versions of implementing the SEP.
3. Assess the relative performance of the SEP in two major and different STEM education contexts.
5. Enhance evaluation cyberinfrastructure (the Netway) through the development and testing of researcher and policy-maker portals.
6. Train the next generation of STEM education evaluators and enhance evaluation capacity within STEM education generally.

D. Project Design

Project activities will be divided into several distinct subprojects, each of which will be conducted by multidisciplinary teams that include educators, scientists, evaluators, software engineers and graduate and undergraduate assistants:

- Virtual SEP Development
- Phase II Study
- Evaluation Cyberinfrastructure Development

Virtual SEP Development

By August 2008, the original NSF grant will be complete and one of its major products will be a facilitator’s guide to the Systems Evaluation Protocol (SEP) (see http://evaluation.cce.cornell.edu/sep/DraftSEPProtocol.doc for a draft of the SEP). This guide is intended to be used by a trained evaluator who would facilitate the implementation of the SEP in a partnering organization. The virtual SEP will require considerable redesign of this original model so that it can be implemented by an organization as a self-guided approach. The virtual SEP will leverage technological resources (i.e. the internet, podcasts, online presentations) in order to make SEP accessible to educators and site-level staff who are not trained in evaluation.

In the first year, the focus of the work will be on developing downloadable materials and presentations based on the original version of the SEP. These tools will be pilot tested and adjustments will be made to the Virtual SEP in year two. The Virtual SEP and the original SEP will then be implemented in similar organizations and the resulting Evaluation Plans will be compared.

Phase II Trial of the Systems Evaluation Protocol

The core of this project consists of the Phase II trial of the SEP. This will be accomplished using four overlapping cohorts of participant sites. The first cohort consists of the seven Cornell Cooperative Extension (CCE) sites and the Cornell Center for Materials Research (a MRSEC site) that participated in the Phase I trial of the Systems Evaluation Protocol (pilots). The second cohort (early adopters) consists of three additional CCE sites and three additional MRSEC sites. The third cohort (middle adopters)
consists of an additional five CCE sites and five MRSEC sites. The fourth cohort (later adopters) consists of five more CCE sites and five more MRSEC sites.

In year one, we will evaluate the effectiveness of the implementation of the evaluation plans completed by the initial pilot sites, implement the facilitated SEP with the second cohort (early adopters), work with the “early adopter” sites to develop the Virtual SEP, and recruit five MRSEC sites and five CCE sites who will comprise the “middle adopters” cohort. In year two, we will test and revise the Virtual SEP, collect baseline data from the “middle adopters”, evaluate the effectiveness of the implementation of the evaluation plans completed by the initial pilot sites and the “early adopters”, and recruit five additional MRSEC sites and five additional CCE sites who will comprise the “later adopters” cohort. In year three, we will implement the Virtual SEP with the “middle adopters”, collect baseline data from the “later adopters”, and evaluate the effectiveness of the implementation of the evaluation plans completed by the initial pilot sites, the “early adopters”, and the “middle adopters”. In year five, we will continue to monitor and assess the effectiveness of the implementation of the evaluation plans for all cohorts and conduct data analyses.

**Research Questions**
The primary research questions that will be addressed are:

1. To what degree is the SEP associated with changes in organizational evaluation capacity and performance outcomes?

2. How do alternative versions of SEP implementation (facilitated versus virtual) differ in performance?

3. How does the relative performance of the SEP differ between MRSEC and CCE sites?

4. To what degree is there cross-organizational adoption of the SEP?

5. How effectively do participants utilize and implement the primary components of the SEP (e.g. developing logic models, pathway models, life cycles, and evaluation plans)?

6. To what degree do participants utilize the SEP’s cyberinfrastructure (i.e the Netway)?

**Participants**
Participants will include the NSF Materials Research Science and Engineering Centers (MRSECs) and extension programs in New York State associated with the National 4-H Science, Engineering, and Technology (SET) goal. In the Phase I trial of the SEP, we partnered with extension offices in six counties as well as New York City and with the Cornell Center for Materials Research (CCMR), a MRSEC site. These partnerships will build upon and expand the project team’s current relationships with the Cornell Center for Materials Research (CCMR) and Cornell Cooperative Extension (CCE). During Phase I, these organizations were an integral part of creating the SEP, serving as both users and advisors to create a user-friendly and practical evaluation system. In addition to our continuing work with these offices, we will also recruit a total of 13 additional CCE offices and 13 additional MRSECs to participate in the Phase II trial of the SEP.

The first cohort of participants (pilots) includes the seven CCE pilot sites and the Cornell Center for Materials Research. We will continue to collect data from these sites in the Phase II trial of the SEP. Data from this cohort will contribute to assessing how effectively participants utilize and implement the primary components of the facilitated SEP (research question 5). The second cohort of participants includes early adopters of the facilitated SEP in both the CCE and MRSEC systems. Three additional MRSEC sites (University of Chicago Materials Research Center, Columbia University Center for Nanostructured Materials, The University of Southern Mississippi Center for Response-Driven Polymeric Films) will be added in Year one of the REESE project and will receive the facilitated SEP. Three additional CCE program sites involved in the 4-H SET initiative will also be added in Year one and will receive the facilitated SEP. In Year three, five MRSEC sites and five CCE offices engaged in the 4-H
SET initiative will be invited to participate in the Phase II Trial of the SEP. In Year four, an additional five MRSEC sites and five CCE sites will be added. A total of ten program sites from the MRSEC system and ten program sites from the CCE system will be selected to participate based on a competitive application process. All selected sites (middle adopters and later adopters) will receive the Virtual SEP. The selection of these sites will provide both the comparative data needed to assess differences between the Facilitated SEP and the Virtual SEP (research question 2) and will also provide data on the cross-organizational adoption of the SEP (research question 4).

**Measures**

A variety of measures and observations will be included in the assessment of the SEP, including:

1. Evaluation capacity checklists (for both the organizations and their participating programs)
2. Logic model rubric; this is part of blind-review to assess the relative strength of the logic model.
3. Evaluation plan rubric; this is part of a blind-review to assess the relative strength of the evaluation plan.
4. Netway usage statistics (e.g. overall activity, usage by site, usage by program area, usage over time).
5. Lifecycle analysis; this is a component of the SEP. The goal of the systems approach used in the SEP is for programs to progress in their Lifecycle phase. Therefore, a measure of success is that programs do not remain static, but rather progress to a new stage of both the Program Lifecycle and the Evaluation Lifecycle. Additionally, the SEP also encourages alignment of Program Lifecycle and Evaluation Lifecycle. Lifecycle Analysis will be completed both as part of the SEP and as a measure of effectiveness.
6. Evaluation plan review and assessment; this tool will be developed in Year one and will be used to assess effective implementation of the Evaluation Plan.

**Design**

This study will be conducted using a longitudinal-sequential (or cohort-sequential) design with four cohorts over five years (Figure 2). The first cohort is comprised of the eight pilot sites (seven CCE, one MRSEC) already engaged in the SEP process. Data from the initiation stage of this process (pilot years 1 and 2) have already been collected. Over the next five years we will continue to follow these pilot sites through their implementation stage (pilot years 3-7) in an effort to determine the degree to which any changes in evaluation capacity are sustained over time. The second cohort will include six additional sites who will engage in the initiation stage of the SEP process (REESE years 1-5). In addition, we will monitor the evaluation activity of five CCE sites and five MRSEC sites who have not participated in the SEP process in REESE year 2 as a control group.

<table>
<thead>
<tr>
<th>Group</th>
<th>Pilot Year</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cohort 1: Pilots</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Facilitated SEP</td>
</tr>
<tr>
<td>CCMR (n = 1)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Facilitated SEP</td>
</tr>
<tr>
<td>CCE (n = 7)</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Facilitated SEP</td>
</tr>
<tr>
<td>Cohort 2: REESE Early Adopters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Facilitated SEP</td>
</tr>
<tr>
<td>MRSEC (n = 3)</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Facilitated SEP</td>
</tr>
<tr>
<td>CCE (n = 3)</td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td>Facilitated SEP</td>
</tr>
<tr>
<td>Cohort 3: REESE Middle Adopters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Virtual SEP</td>
</tr>
<tr>
<td>MRSEC (n = 5)</td>
<td>x*</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Virtual SEP</td>
</tr>
<tr>
<td>CCE (n = 5)</td>
<td>x*</td>
<td>x</td>
<td>x</td>
<td>x</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Virtual SEP</td>
</tr>
<tr>
<td>Cohort 4: REESE Later Adopters</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Virtual SEP</td>
</tr>
<tr>
<td>MRSEC (n = 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>Virtual SEP</td>
</tr>
<tr>
<td>CCE (n = 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>x</td>
<td>x</td>
<td></td>
<td>Virtual SEP</td>
</tr>
</tbody>
</table>

*x* Data will be collected from the programs and serve as comparison groups for the REESE early adopters and middle adopters.

Figure 2. Study design.
**Analyses**

The longitudinal-sequential design will allow us to make a number of different comparisons both between and within the four cohorts. Primarily, we will use change score analysis to determine a) if there are any changes over time that are associated with exposure to the SEP process, b) if these changes differ between groups, and c) if these changes are sustained over time.

Each research question will be addressed as follows:

**Research Question 1:** To what degree is the SEP associated with changes in organizational evaluation capacity and performance outcomes?

All of the participant sites will complete pre/post evaluation capacity assessments every year that they are involved in the study. Within group change scores will be assessed using the Evaluation Capacity Checklists. Five “middle adopter” sites from CCE and five “middle adopter” sites from MRSEC will complete pre/post evaluation capacity assessments at REESE year two, however none of the “middle adopter” sites will receive treatments until REESE year three. These year two pre/post assessments will serve as a control comparison for the REESE “early adopter” sites. Similarly, the five “later adopter” sites from CCE and five “later adopter” sites from MRSEC will complete pre/post evaluation capacity assessments at REESE year three and will not receive treatment until REESE year four. The “later adopters” will serve as a control comparison for the “middle adopter” sites.

**Research Question 2:** How do alternative versions of SEP implementation (facilitated versus virtual) differ in performance?

Comparisons will be made between the six “early adopter” sites (three MRSEC, three CCE) all of whom will receive the Facilitated SEP and the twenty “later adopter” sites (ten MRSEC, ten CCE) all of whom will receive the Virtual SEP. Pre/post change scores on the Evaluation Capacity checklists, post-only scores on the Logic Model Rubric, post-only scores on the Evaluation Plan Rubric, and Netway usage statistics will be compared.

**Research Question 3:** How does the relative performance of the SEP differ between MRSEC and CCE sites?

Comparisons will be made between the “early adopter” MRSEC sites and the “early adopter” CCE sites on all measures. Comparisons will also be made between the “middle adopter” MRSEC sites and the “middle adopter” CCE sites on all measures. Comparisons will also be made between the “later adopter” MRSEC sites and the “later adopter” CCE sites. Pre/post change scores on the Evaluation Capacity checklists, post-only scores on the Logic Model Rubric, post-only scores on the Evaluation Plan Rubric, and Netway usage statistics will be compared.

**Research Question 4:** To what degree is there cross-organizational adoption of the SEP?

Not all of the eligible programs will be able to use the SEP, therefore, one measure of cross-organizational uptake will be a comparison of the number of programs offered access to the SEP and the number of programs that actually complete the SEP. Comparisons of cross-organizational adoption will be made between the MRSEC system and the CCE system. Netway usage statistics will be collected and compared as will the use of technical assistance resources available in conjunction with the Virtual SEP.

**Research Question 5:** How effectively do participants utilize and implement the primary components of the SEP (e.g. developing logic models, pathway models, life cycles, and evaluation plans)?

Once program sites have completed their first year of evaluation planning (with either the Facilitated or Virtual SEP), follow-up assessments will be conducted to determine how effectively the Evaluation Plans were implemented. In particular, we will assess the following: Were all components of the Evaluation Plan completed; Do revisions of the initial Evaluation Plan in the second (and third, fourth, etc.) years of implementation reflect feedback received in previous years’ evaluation results; Do evaluated programs move to a different phase of the Lifecycle? Netway usage statistics will also be analyzed.

**Research Question 6:** To what degree do participants utilize the SEP’s cyberinfrastructure (i.e the Netway)?

Netway usage statistics for all groups at all years of the study will be assessed.
Schedule and Timetable
Following is a graphical representation of the proposed study schedule and timetable for the four cohorts over the five-year project (Figure 3).

![Figure 3. Study schedule and timetable.]

Evaluation Cyberinfrastructure Development
An initial version of the Netway cyberinfrastructure was developed by the P.I. as part of the Phase I project that is currently being completed. It is being beta-tested at NSF-funded STEM programs at the Cornell Center for Materials Research and at the Museum of the Earth and in the Cornell Cooperative Extension system in six counties in New York State (Chenango, Jefferson, Onondaga, St. Lawrence, Tompkins, and Ulster) and in New York City; some of these extension programs are STEM education programs, and programs in all counties include youth development. The software currently enables an educator to enter information about an educational program (inputs, assumptions, contextual issues, activities, outputs and outcomes) and its evaluation (questions, participants, measures, design, analysis, reporting) and immediately generate both a logic model and a draft evaluation plan. In addition, the system can be used to enter program participant data and produce simple reports about participation by activity, program or program area. This proposal includes funding to accomplish the adaptation and enhancement of the existing Netway software required for this research.

Virtual Systems Evaluation Protocol Support. The Netway is currently constructed to be used by a trained facilitator. Under the proposed work, it will be significantly enhanced by developing the required user interfaces and data structures to enable self-guided use of the Systems Evaluation Protocol, in effect creating a Virtual SEP or vSEP. This will require the development team to conceptualize, design, implement and test such a capability. Several major changes will be needed to the user interface of the existing system. Most notably, the current help function will need to be redesigned to provide a much more extensive and refinable interface that can be easily edited by the development team and can provide...
interactive forms and wizards that the user can be guided through to complete the necessary steps in the protocol. This essentially requires that the current passive help system where the user must ask for help be transformed into an active function that anticipates the user’s place in the protocol and provides necessary guidance. This will also require the integration of the current manual SEP exercises and forms into the system.

Measurement Systems Integration. The current Systems Evaluation Protocol involves the collection of a broad range of measures including

- Organizational Evaluation capacity checklist to assess current organizational capability for evaluation.
- Program Evaluation Capacity checklist to assess current program capability for evaluation.
- Logic model rubric to assess the relative strength of the program logic model.
- Evaluation plan rubric to assess the relative quality of the evaluation plan.
- Evaluation plan implementation assessment to assess effective implementation of the Evaluation Plan.
- Program Lifecycle to assess program lifecycle phase.
- Evaluation Lifecycle to assess evaluation lifecycle phase.
- Netway usage statistics (e.g. overall activity, usage by site, usage by program area, usage over time).

Several of these (lifecycles and Netway usage) have been partially implemented in the current system. We propose significantly enhancing the measurement capabilities of the Netway by integrating multi-user data entry of all measures over time and providing an integrated evaluation cyberinfrastructure that can be used to support this NSF proposal. In addition, we propose to build basic analytic capabilities that will enable reporting of all data at all levels (Netway, system, office, program, project, activity) and over time (e.g., multiple waves of measures to assess change), and to develop comprehensive export facilities to enable extra-Netway additional statistical analysis of results.

Because we are the original developers of the software, we have the programming design expertise already in place, and are familiar with the existing data structures. Approval of this funding will enable us to adapt the software on an ongoing basis and incorporate it into the Phase II trial. The software will provide an important tool that will be essential for supporting both implementations of the SEP protocol, and especially for the virtual self-guided implementation.

E. Project Management

The project will be led by Dr. William Trochim, PI (Cornell University). Day to day operations for the project will be overseen by Dr. Trochim with the assistance of a Project Manager to be named. This team will guide the other personnel, namely graduate students and undergraduates, to completion of the project goals. The graduate and undergraduate students will be recruited specifically through Historically Black Colleges and Universities (HBCUs) and Quality Education for Minorities (QEM) as well as through traditional recruiting activities such as presentations by our faculty and students at national meetings.

F. Timeline

In Year 1, the project team will build the Netway cyberinfrastructure to support its implementation on a national level. This will include creating a self-guided, Virtual Systems Evaluation Protocol. Graduate students will be recruited to synthesize research on relevant evaluation methods and analysis procedures. In Year 1, implementation of the evaluation plans completed by the initial pilot sites will be evaluated and the facilitated SEP implemented with the second cohort (early adopters). Early adopters will be engaged in development of the Virtual SEP and ten sites recruited for Year 2. Results from Year 1 will be shared at national conferences.

In Year 2, the Virtual SEP will be tested with the middle adopters and the effectiveness of the implementation of the evaluation plans to date evaluated. Ten additional sites will be recruited for Year 3.
(later adopters) This stage will involve testing the Virtual Systems Evaluation Protocol and Netway and refining them to better meet the needs of the STEM practitioners in the partner organizations. Graduate students will be intimately involved in this research, gaining both evaluative research skills and communication skills. Preliminary results from both research and implementation will be shared at national conferences and through peer reviewed publications.

In Year 3, the Virtual SEP will be implemented with the middle adopters and baseline data collected from the later adopters. Effectiveness of evaluation plans completed by the sites to date will be evaluated. The Virtual Systems Evaluation Protocol and Netway portals will continue to be refined. Lessons learned in both development and implementation of the Protocol will be shared at national conference presentations and in peer reviewed publications.

In Year 4, the Virtual SEP will be implemented with the later adopters. Effectiveness of the implementation of evaluation plans completed by all sites to date will be evaluated.

In Year 5, the effectiveness of implementation of the evaluation plans for all sites will be evaluated, and data will be aggregated and analyzed across all cohorts. The Virtual Systems Evaluation Protocol and Netway will refined based on results to date, with the end product being a fully functioning, many-tiered portal for STEM education evaluation. Results from ongoing research and the project overall will be shared at national conferences and in peer reviewed publications.

**G. Evaluation**

The evaluation of this project will utilize the same tools and measures being used to assess effective implementation of the SEP in the MRSEC and CCE sites. We developed a logic model, pathway model, and evaluation plan for the SEP during the Phase I implementation of this project. These will be used as baseline measures in Phase II. Our overarching goal is to move the SEP to Phase II in its programmatic and evaluation lifecycles. The degree to which we are successfully able to implement the SEP as a Phase II program will be assessed by the dissemination of the SEP to multiple program sites, the development and implementation of the Virtual SEP, and the degree to which CCE and MRSEC sites are able to move their projects further along in their Lifecycle phases over time.

**H. Key Staff, Consultants, and Advisors**

**Project Team**

**William Trochim, P.I.**

The P.I. for this project will be Prof. William Trochim, Ph.D., a professor in Policy Analysis and Management, the Director of Evaluation for Extension and Outreach at Cornell, the Director of Evaluation of the NIH Weill Cornell Clinical and Translational Science Center, and the Director of the Cornell Office for Research on Evaluation (CORE). Prof. Trochim is one of the leading evaluation experts in the world and is uniquely qualified to guide this project. A graduate of the Ph.D. program in Methodology and Evaluation Research at Northwestern University, Dr. Trochim has taught evaluation and research methods at both the undergraduate and graduate levels at Cornell since 1980 and has mentored over fifty Ph.D. students in evaluation. Dr. Trochim is one of the leaders in the emerging area of evaluating large research center grant initiatives, especially within the NIH and NSF research systems. Dr. Trochim also was a core member of a special NCI effort on systems thinking and modeling called the Initiative on the Study and Implementation of Systems (ISIS) project, is a co-author of an in press monograph on this project, and is lead author of an article on challenges of systems thinking and modeling in public health in a special issue of the American Journal of Public Health in March, 2006. In his role as Director of Evaluation for Extension and Outreach, Dr. Trochim has developed the Systems Evaluation Protocol and implemented it in over 50 education and outreach programs, including a number of NSF-funded STEM programs. Dr. Trochim is the President of the American Evaluation Association, the leading professional organization in the field of evaluation.
Jennifer Brown, co-Investigator, will be an assistant professor at a to-be-determined university. She has been the graduate research assistant on the Systems Evaluation project for two years. Her work has focused on developing innovative ways to integrate research and practice and on the development of the SEP and related materials. Her research also focuses on understanding the relationship between involvement in youth development programs and positive youth development.

I. Dissemination

The project team will train the next generation of STEM evaluation graduate students who will move on to professional positions as ambassadors for the project. Research results will be shared primarily through conference presentations and peer reviewed publications. A web site will be created that shows partners, research updates, and other information relevant to the project.
References


Capra, F., Mill Valley School District (Mill Valley Calif.), & Elmwood Institute. (1994). From the parts to the whole: systems thinking in ecology and education. S.l.: s.n.;


Appendix: Coordination Plan

Coordination of Project Teams. Four separate and interconnected interdisciplinary collaborative teams will be created to achieve the proposed activities. Each team is organized conceptually around one of the three major activities (with separate teams for implementing the SEP in the two research contexts):

- **Virtual SEP Development Team.** This team will be led by J. Brown, supported at her site by a graduate RA and by the Project P.I., Project Manager and graduate RA at Cornell. They will be responsible for the development of the Virtual SEP protocol including manuals and online materials.

- **Phase II Study MRSEC Team.** This team will be led by the CCMR Research Assistant with support from the project P.I., the Project Manager, the CORE RA and a MRSEC advisory group consisting of the MRSEC evaluation network SEP participants. This team will be responsible for management of all aspects of the MRSEC multiple cohort SEP study.

- **Phase II Study CCE Team.** This team will be led by the Manager of Evaluation for Extension and Outreach with support from the project P.I., the Project Manager, the CORE RA, a national Cooperative Extension Advisory Group and the CCE evaluation network that includes SEP Evaluation Project Managers from new project sites and previous pilot sites. This team will be responsible for management of all aspects of the CCE multiple cohort SEP study.

- **Evaluation Cyberinfrastructure Development Team.** This team will be led by the P.I. and will include the Project Manager, CORE graduate RA, an undergraduate assistant with expertise in web development and the contracted programming team. This team will be responsible for all development, maintenance and support of the Netway system for this research.

The work of the four teams on this research project will be coordinated through the Cornell Office for Research on Evaluation (CORE) which is directed by the P.I. (Trochim). CORE has physical locations on both the Ithaca campus of Cornell University and at the Weill Cornell Medical College in Manhattan, and offers university-based support for the development and implementation of internal and external evaluations. The Ithaca campus office supports grants from the National Science Foundation, the National Institutes of Health, the State of New York, and others. The Manhattan CORE office serves the NIH-funded Clinical and Translational Science Center at Weill Cornell Medical College. CORE provides consultation and collaboration on evaluation in connection with funded research grants, is involved in the development and testing of evaluation methods, measures, tools and systems; coordinates resources and expertise from throughout the evaluation field; and, provides direct consultation on evaluation pilot and development projects on both Cornell campuses and at educational extension and outreach sites throughout New York State. Finally, CORE facilitates both communication and education about evaluation systems and the dissemination of evaluation results through its cross-campus and statewide evaluation networks and through peer-reviewed scientific research publications.

The project teams will meet separately on a regular basis and will convene annual by teleconference (with Adobe Connect for video and computer connections) on a semi-annual basis to coordinate cross-team integration.

Data Coordination and Integration. The primary data collection for this project will be coordinated through the Netway cyberinfrastructure system. The Netway enables the construction of any number of systems, offices, program areas and projects in a multiple hierarchies. In this study, separate systems will be constructed for the MRSEC and CCE contexts, with each system having one office per site. Within each site it will be possible to construct one or more program areas, each of which may contain one or more projects. For example, Figure A.1 shows a sample Netway screen shot that lists multiple programs (i.e., projects) in the general program area of “Family and Youth Development” in the New York City office of the current CCE system.
Within each project, the Netway gathers considerable data regarding program logic and pathway models. For program models data is collected on the following: name, description, mission, program lifecycle, assumptions, context, inputs, activities and outputs/outcomes. The program automatically generates traditional columnar logic models and enables interactive construction of pathway logic models. In a traditional logic model, program activities, outputs and outcomes are listed in separate columns but are not individually linked to one another. This is in contrast to models developed under Theory of Change approaches where causal “pathway” models are the more typical representation. While similar in general to columnar logic models, the pathway model explicitly links individual activities, outputs and outcomes with causal arrows. Examples of screen shots of these two approaches in the current Netway are illustrated in Figure A.2.

When we have multiple educational programs – a “system” of programs – as is the case in STEM education some of the programs are likely to have identical activities, outputs or outcomes. For instance, two STEM education programs might share a common activity like a hands-on lesson with a microscope. Or, they might both have a mid-term outcome of affecting children’s attitudes towards science. When we have causal pathway models that share elements, we term this a **networked pathway** model which is what gave rise to the name for the software. A netway model is a representation of a system of programs. In the current version we developed data structures that enable us to identify and depict netway models of
this nature. Figure A.3 shows a visual representation of the netway data structure across multiple program pathway models.

![Netway Data Structure Diagram]

Figure 3. Hypothetical case of four program pathway models that have elements in common.

For evaluation plans, the Netway enables entry and storage of the following data: evaluation questions, sample, measurement, design, analysis, reporting timeline and evaluation cycle.

The Netway will also be used to store data from all of the measures collected for the Phase II SEP study, including: organizational and program evaluation capacity checklists; the logic model review assessments; the evaluation plan review assessments; the lifecycle analyses; and the evaluation implementation assessments.

In addition, the Netway automatically stores comprehensive user statistics including log data on every user, entry and report. This will enable us to do extensive analysis of Netway use over time in order to track the rates at which logic models and evaluation plans are constructed and revised and to link these data with users, projects, programs, offices and systems.

The Netway is secure (SSL) web-based system with separate user identities and multiple status levels that allows total control over different access levels. Each system user is assigned a userid and password that is unique and is required to log onto the system. For instance, a site coordinator can be assigned full editing and review permissions for all programs and projects within their office, a staff member responsible for a specific project can be limited to only editing that project’s information, and an evaluation consultant can be assigned privileges for reviewing and assessing all of the projects and programs within an office or system.

The Netway system will be used as the primary mechanism for data collection and reporting for this research project. It currently enables extensive reporting on project logic models, evaluation plans, usage data, and project and evaluation lifecycles. In addition, the Netway has extensive data export functions that enable system administrators to export data at any level (system, office, program, project).

One unique feature of the Netway is that it enables anyone in the system to view (but not edit) and access project logic model and evaluation planning information anywhere else in the system. Whenever an educator enters information about a program activity, output or outcome, the software does a simple dynamic search for other programs already in the system that already have similar entries. These are displayed for the user as options they could select. For example, if an educator enters an activity like “hands-on demonstration with a microscope” the software shows any other existing entries that use those key terms, with the best fitting ones shown first. The educator has the option of choosing any existing
entities that they think appropriate, or of continuing to enter their own. In addition, educators can view reports of the logic models and evaluation plans of other projects or programs. This feature has several important effects from a systems perspective. It helps the educator think of things they might not have thought of on their own. It shows the educator that there are others out there doing this kind of activity or looking at this kind of output or outcome. And, if they select an existing entry, it immediately links their pathway model to another, while creating a new node on the Netway. Note that if they elect not to choose an existing element, the one they enter becomes part of the system and may be selected subsequently by others. Once a connection is made to another pathway model, it is possible for the educator to benefit by seeing the other pathway linked to the connected element. Continuing with the microscope example, if the educator selects an existing activity that is identically or similarly worded to “hands-on demonstration with a microscope”, the software would also be able to display the programs, activities, outputs and outcomes that were connected with that activity in the other pathway model. This is analogous to the function in many current websites like amazon.com that show “others who made this selection also liked…” when you select a book or movie.

This is a simple and profound change in traditional database structure. It is inherently a systems approach in that it connects local-local nodes and enables a global network to emerge from these connections organically. Over time, and with more users, the web-based system becomes more powerful and useful both as a data collection mechanism and as a vehicle for cross-program interaction. The probability that an educator/user would find features that match what they are interested in doing would increase. The system enables local-local learning to occur immediately at the point of need for the educator. They immediately see other relevant information that can be used to improve their program planning and evaluation. And, in the major STEM education topical areas, it is likely that over time a broad-based consensus would emerge where new programs that replicate older ones would find that they could quite rapidly adopt the standards used earlier. At the same time, the system allows for continuous adaptability, even in more established, emergent, normative areas.

To assure that there is consensus and mutual understanding regarding data access and availability the project teams will negotiate a section on data access into the written Memorandum of Understanding for the team. The results of these negotiations will be integrated into the comprehensive data use plans and included as part of the IRB proposal for this project.

**Communication and Publication Plans.** A key feature of the Coordination Plan will be the negotiation of written Memoranda of Understanding (MOU). A separate MOU is already built into the SEP process and will be negotiated with each MRSEC and CCE site. This MOU details the roles and responsibilities of the site participants and of evaluation team members. In addition, a separate MOU will be developed for each of the four project teams that will detail their roles and responsibilities, including plans for communicating throughout the project, access to project data, and publication rights and responsibilities. Finally, a cross-team integrated MOU will be developed that will detail relevant cross-team roles and responsibilities, especially with regards to data access and analysis and to presentations and publications. Subject to a final negotiated agreement, it is expected that each of the teams will have rights to analyze and present or publish data that is relevant to their primary team responsibilities. For instance, the virtual SEP development team is expected to write about the development of this variation on the protocol and on analyses of Netway data relevant to its implementation. Both the MRSEC and CCE teams will be expected to develop publications about the use of the SEP in their respective areas (materials research and engineering education for the MRSECs; 4-H youth development for the CCEs). In addition, it is expected that there will be multiple cross project and cross-team publications that will describe the results of this research to STEM education, evaluation and general audiences.
December 12, 2007

Professor Bill Trochim
249 MVR
Cornell University
Ithaca, NY 14853

Dear Bill:

I am writing to share our enthusiasm for continuing to work with you on improving the quality of STEM evaluation for New York State 4-H. We know from our experiences that 4-H changes lives, helping youth to become confident, mature adults ready to succeed in today’s challenging world. Studies show that youth participating in 4-H do better in school, are more motivated to help others, feel safe to try new things, and develop lasting friendships. In 2006, some 458,000 youth from urban, suburban, and rural communities across New York participated in 4-H experiences. Many of these youth participated in Science, Engineering and Technology programs through projects such as rocketry, GPS mapping, and DNA analysis.

We have an active Science, Engineering and Technology program in our state, and it is difficult to evaluate these programs which are many and varied. Each county implements its programs in a unique way that meets the needs of its participants, so it is always a challenge to create flexible evaluation resources to guide our extension agents and program staff.

As you know, in the past two years, we have had the opportunity to work with you on evaluation of STEM programs in New York City and six counties, with over fifty programs benefiting from the work that was undertaken. This work has had a tremendous impact on how program staff approaches evaluation and on the implementation of evaluations at each office. The Evaluation Planning Partnerships have improved the nature and impact of evaluation for our programs, as evidenced by the outcomes shared at our statewide meeting the first week of December, 2007. Members of our program staff have very positive things to say about how things are progressing, and the rapport that you have with each office is excellent.
We are happy to participate in future work by expanding this pilot effort to all of the counties in the state. Our 4-H Youth Development program is strong, and our staff will welcome the opportunity to work with your team on expanding the evaluation resources to their colleagues.

Please let me know if you have any questions or need any further information about our programs. Thank you for continuing this valuable project into the future.

Sincerely,

Barbara Schirmer
Assistant Director, Cornell Cooperative Extension
State Program Leader, 4-H Youth Development
Prof. William Trochim  
Cornell University  
249 MVR  
Ithaca NY 14853

December 12, 2007

Re: Support for REESE Proposal

Dear Bill:

As Director of the Cornell Center for Materials Research, a NSF-funded Materials Research Science and Engineering Center (MRSEC), I am writing to express our sincere interest in continuing to work with you on improving the quality of STEM evaluation for our educational programs. Our center has a vibrant K-12 education and outreach program. In 2006 alone, CCMR programs reached more that 2500 K-12 students, 350 parents and 340 K-12 teachers through more than 60 unique educational modules. Evaluation and assessment of our programs are crucial to further development, and we are always seeking new ways to evaluate our portfolio of projects. We have already developed many new methods and tools with your team and are eager to continue this process.

In addition to enriching our own programs, we are excited by the possibility of sharing and disseminating the new assessment tools and methods with other MRSEC programs around the country. The NSF currently funds 29 MRSEC programs across the country, and all have substantial K-12 educational outreach programs. Like us, the other centers also have a growing need for efficient, focused evaluation of their education and outreach programs. Many of the other centers have expressed an interest in using the resources that you have developed. The REESE grant would provide the resources necessary for this task.

Over the course of the proposed project, we will plan to be involved with your program in a number of ways. First, we will work with you on the various stages of implementation of the Systems Evaluation Protocol, including forming partnerships between the Systems Evaluation team at Cornell and the partnering sites (Evaluation Planning Partnerships) and implementing the systems evaluation process (Protocol). Second, we will help you form connections to a small subset of interested MRSEC sites; this subset will help us refine our program and help develop dissemination tools. We have initially targeted the large, well established MRSEC programs at the University of Chicago, Columbia University and the University of Maryland at College Park for this phase. Using the knowledge gained from these small-scale interactions, we will then work with your team to broadly disseminate your tools and assessment knowledge to all MRSEC sites as well as other NSF-funded centers (e.g., NSECs, STCs).

Our involvement in this project will require 50% effort of an educational outreach staff member, in part to ensure that we effectively engage and assist the partner MRSECs. Over the course of the five-year grant, we anticipate working with the majority of the other MRSEC sites (i.e. more than 20 sites), forming communication linkages that will support the goals of the grant as well as our ongoing collaborations with these centers.
Please do not hesitate to contact me if I can provide further information or assistance in support of your project. We very much look forward to the successful implementation of this project and a long and fruitful collaboration improving the evaluation infrastructure of the MRSEC program.

Sincerely,

Melissa A. Hines
Director, Cornell Center for Materials Research
Professor of Chemistry
December 19, 2007

Professor Bill Trochim
249 MVR
Cornell University
Ithaca, NY 14853

Dear Bill:

I am writing to express our interest in continuing our collaboration with you to improve the quality of evaluation of our extension educational programs, and particularly in relation to youth development programs related to the national Science, Engineering and Technology (SET) goal. As you know, Cornell University is the Land Grant College for the State of New York with comprehensive coverage of the state through our county-based Cornell Cooperative Extension (CCE) Associations and the New York City office that provide 56 separate local sites in our system. We are positioned to make a positive impact on the lives of young people by engaging them in our Science, Engineering, and Technology program offerings. Our New York State 4-H youth development ranks third in the country in state 4-H enrollment, with more than 316,000 youths, 5 to 19 years old, participating. This is an exceptional resource and one that will provide an excellent venue for assessing the efficacy of the Systems Evaluation Protocol that you have developed.

We are fortunate to have become involved in your work to create a new model for STEM program evaluation. This initial work with six counties and New York City led to the production of logic models and evaluation plans for over 50 programs, and significantly enhanced evaluation capacity within our extension offices. The partnerships with you and your team have been effective in engaging extension personnel to think broadly about the goals and outcomes of the programs that they offer, while also providing useful tools to evaluate specific activities.

As part of your newly proposed work, the State 4-H Youth Development program will work closely with you and our county associations throughout New York State to invite them to participate in this rigorous test of the efficacy of your model. The accessibility and ease of use of the online "Netway" system will greatly enhance the connections between Cornell and the individual counties. From a management perspective, it will be incredibly useful for me and CCE to be able to participate as users.
of the "Netway" at any point and to see the evaluation plans, logic models, and pathway models as the counties are creating them. This portal will be an excellent window into the evolution of each county’s evaluation planning and implementation process.

We look forward to working with you on this project. Please do not hesitate to contact me if you have any questions or need additional information.

Sincerely,

Helene R. Dillard, Ph.D.
Director, Cornell Cooperative Extension
Associate Dean, Colleges of Agriculture and Life Sciences and Human Ecology
Professor, Plant Pathology
Dear Bill,

I am writing to confirm my interest in the proposed “Large Empirical Research” project. I am eager to continue our work in STEM evaluation and look forward to the opportunity to expand on the work that I began as your Graduate Research Assistant on your current NSF grant. Expanding on this work will produce both interesting research, as well as, useful tools for informing the field of STEM evaluation.

My past involvement in the development and testing of the Systems Evaluation Protocol will allow me to seamlessly integrate this approach in our future work. The institutional knowledge that I bring to this project will allow me to transition to this new phase of work with relative ease. As a new assistant professor, I will be eager to initiate my program of research and engage younger scholars in the field of evaluation, STEM, and youth development. I am pleased to report that I have already received two job offers for tenure-track faculty positions. As a junior colleague and co-PI on this project, this opportunity exemplifies our efforts to include young scientists (from undergraduates to graduate students, and emerging scholars) in the evolving field of STEM evaluation.

In addition to my knowledge of evaluation in STEM contexts, I have also developed expertise in youth development. My substantive research interests in youth development programs and policies will contribute to our work with programs that target youth populations.

As part of my commitment to this work, I will recruit a graduate student who will be responsible for assisting with work on this project. This student will primarily assist with developing systems for supporting evaluations, synthesizing research on relevant evaluation methods and measures, and collaborating on evaluation pilot studies. Based on the finding from this project, we will present results at national meetings and publish in appropriate journals.

I am very excited about the potential advances that this project can produce and I look forward to working with you. Please contact me if you need any additional information.

Sincerely,

Jennifer Southwick Brown
Professor William Trochim  
Dept. of Policy Analysis and Management  
Room 249 MVR Hall  
Cornell University  
Ithaca, NY 14853  

RE: Research and Evaluation on Education in Science and Engineering (REESE)  

Dear Bill,  

As Director for the University of Chicago Materials Research Center, I am delighted to confirm our plans to collaborate with you on the development of an evaluation protocol for our K-12 Science, Technology, Engineering and Mathematics (STEM) activities.  

As a NSF-DMR Materials Research Science and Engineering Center (MRSEC), we currently offer a broad range of educational programs including middle school science clubs, a four-week math enrichment program for high school students in the summer, teacher professional development workshops leading to IL middle school science endorsement, several museum programs, and a number of public lectures designed to bring the excitement of research to our local inner-city community, impacting roughly 139,000 K-12 students per year. We currently evaluate the success of our programs using instruments and methods developed internally. We would greatly appreciate the opportunity to work with your team to build a robust evaluation strategy that is firmly based on research into evaluation theory and methods.  

The proposed research and evaluation program builds upon an initial collaboration between the Cornell Center for Materials Research (CCMR), another NSF-funded MRSEC, with researchers in the Cornell Dept. of Policy Analysis and Management. In the initial collaboration, the CCMR worked with your team to develop evaluation strategies for module-based STEM activities — single-interaction activities.
that engage students for a few hours. The short nature of these programs makes them particularly difficult to evaluate. In the initial stage of the project, the CCMR created a program logic model, evaluation plan and specific instrumentation for determining the effectiveness of the modules presented to the K-12 community.

In the new phase of the interaction, we look forward to building upon the CCMR’s original research and to developing a robust assessment model that can be disseminated broadly. As one of the first remote sites, we look forward to working with you and the CCMR in the refinement and extension of the current evaluation protocol, as well as in the development of the documentation necessary for further dissemination. We are particularly excited about the possibility of developing a cross-campus control group for comparative assessment. During the course of the grant, we envision the assessment program spreading nationally, first to our campus and other remote pilot sites, then to other MRSECs and NSECs, and finally to all interested outreach programs.

Through our coordinated efforts, we look forward to improving K-12 STEM education across the Nation.

Sincerely,

Stein-Freiler Distinguished Service Professor
Director, University of Chicago Materials Research Center
Department of Physics and the James Franck and Enrico Fermi Institutes
Professor William Trochim  
Department of Policy Analysis and Management  
Room 249 MVR Hall  
Cornell University  
Ithaca, NY 14853  

December 19, 2007  

RE: Research and Evaluation on Education in Science and Engineering (REESE)  

Dear Bill,  

As Director for the University of Southern Mississippi’s Center for Response-Driven Polymeric Films and Coatings funded by the National Science Foundation Materials Research Science and Engineering Center (MRSEC) program, I am delighted to confirm our strong desire and plans to collaborate with your program on the development of an evaluation protocol for our K-12 Science, Technology, Engineering and Mathematics (STEM) activities.  

As a NSF-DMR Materials Research Science and Engineering Center (MRSEC), we currently offer a broad range of educational activities including Research Experience for Undergraduates (REU), Research Experience for Teachers (RET), Summer Science Camp, “What is a Polymer?” K-12 Outreach Programs, and High School Student Research Experiences. These efforts collectively impact over five hundred students annually. We currently evaluate the success of our programs using instruments and methods developed internally and are excited about and appreciate the opportunity of working with your team to build a robust evaluation strategy firmly based on research into evaluation theory and methods.  

It is our understanding that the proposed research and evaluation program builds upon an initial collaboration between the Cornell Center for Materials Research (CCMR), another NSF-funded MRSEC, with researchers in the Cornell Department of Policy Analysis and Management. In the initial collaboration, the CCMR worked with your team to develop evaluation strategies for module-based STEM activities — single-interaction activities that engage students for a few hours. The short nature of these programs makes them particularly difficult to effectively evaluate. In the initial stage of the project, the CCMR created a program logic model, evaluation plan, and specific instrumentation for determining the effectiveness of the modules presented to the K-12 community.
In the new phase of the interactions, we look forward to building upon the CCMR’s original research and developing a robust assessment model that can be broadly disseminated for others to follow. As one of the first remote sites, we look forward to working with you and the CCMR team in the refinement and extension of the current evaluation protocol, as well as in the development of the documentation necessary for further dissemination. Of particular importance is the possibility of developing cross-campus control groups for comparative assessment. During the course of the grant, we envision the assessment program spreading nationally, first to our campus and other remote pilot sites as well as to other MRSECs and NSECs across the Nation, and finally to all interested outreach programs. I am pleased to inform you that Dr. Sarah Morgan, Education Director of the Center for Stimuli-Responsive Polymeric Films and Coatings at USM has agreed to work directly with your team in an effort to implement the outcomes of the program at our site.

Through our jointly coordinated efforts, we look forward to improving K-12 STEM education across the Nation.

Sincerely,

Marek W. Urban
Professor of Polymer Science
School of Polymers and High Performance Materials
Director, MRSEC at USM

cc: Dr. Martha Saunders, President
Dr. C. Burge, VP Research and Economic Development
January 4, 2008

To whom it may concern,

As Director for the Columbia University Center for Nanostructured Materials, I am delighted to confirm our plans to collaborate with you on the development of an evaluation protocol for our K-12 Science, Technology, Engineering and Mathematics (STEM) activities.

As an NSF Materials Research Science and Engineering Center (MRSEC), we currently offer a broad range of educational programs including Research Experience for Teachers (RET), which has allowed 18 science high school teachers to do research in the MRSEC each summer since 1999. In the past, we evaluated the success of our programs using instruments and methods developed internally. We would greatly appreciate the opportunity to work with the CCMR team to build a strong evaluation strategy that is firmly based on research into evaluation theory and methods.

We look forward to building upon the CCMR’s original research and to developing a robust assessment model that can be disseminated broadly. As one of the first remote sites, we look forward to working with the CCMR in the refinement and extension of the current evaluation protocol, as well as in the development of the documentation necessary for further dissemination. During the course of the grant, we envision the assessment program spreading to other MRSECs and ultimately to all interested outreach programs across the country.

Through our coordinated efforts with the CCMR, we look forward to improving K-12 STEM education.

Sincerely,

Irving P. Herman
Professor, Applied Physics
Director, Columbia University Materials Research Science & Engineering Center