1. Activities

This research has been grounded within the concepts of systems and systems thinking. Utilizing this theoretical base contributed to our understanding of how regional evaluation systems - such as those used at the university, school or other community levels - might best be developed and integrated. A systems thinking approach was applied in the development of specific measures and tools of relevance for monitoring evaluation capacity. It also contributed to the development of an evaluation system, in this case, one that spans the STEM endeavor.

The goal of this research has been to design a systems evaluation methodology, and to enhance the capability of the field of evaluation to develop more effective evaluation systems using this methodology. In other words, we present here an evaluation system for systems evaluation. An evaluation system refers to the comprehensive and integrated set of capabilities, resources, activities and support mechanisms for conducting evaluation work. This should not be confused with systems evaluation, whereby we refer to assessment of the functions, products, outcomes and impacts of a system (set of programs, activities or interventions).

The specific context for this effort was the evaluation of Science, Technology, Engineering and Mathematics (STEM) education programs, but its implications extend to the more general field of evaluation. The intellectual merit of this research includes its contribution to the improvement of our understanding of the role of systems in evaluation. Specifically, this has been attained through (1) development of measures of capacity of evaluation systems; (2) development of a cyberinfrastructure that promotes cross-system collaborations; and (3) administration of a pilot test for an evaluation capacity-building system and measure(s). The research has resulted in a set of specific products that have immediate utility in STEM education, including a measurement approach and instruments that can be used to assess and monitor evaluation capacity and support in a STEM outreach and educational system.

1.1. Evaluation of Large Research Initiatives

We contend that the primary contemporary challenge to the field of STEM education evaluation is a systems one. STEM education takes place in a great variety of settings associated with both formal and informal educational contexts. It is funded through large research center grants and through museum outreach programs, local youth development organizations and national systems like Cooperative Extension. Many STEM education projects are small, ad hoc efforts that are designed as outreach and dissemination add-ons to larger scientific endeavors. While some STEM education is based on well-researched educational methods and processes, much of it is not. The system, to the extent that it can be called that, is often disjointed, unconnected, non-cumulative and is generally not mutually informing.

The central evaluation problem is how best to make a (causal) connection between the many and varied local STEM activities and more macro longer-term global outcomes. This is a classic systems thinking problem, a part-whole or “local/global” challenge that every local STEM education project faces – how do we connect the many varied local STEM experiences with the broader global outcomes of interest? Local STEM education projects are typically most concerned with the experiences of their program participants and how service delivery can be improved (local level concerns). The research community more typically works on more global questions related to longer-term impact and the development of a
STEM educated workforce for the future (global level concerns). Connecting these local and global concerns requires more effective integration of the practice and research communities that focus on them.

The current research begins to address the need for integrating research and practice so that practitioners have the tools needed to evaluate their short-term outcomes as well as the ability to tap into the research evidence-base that allows them to make logical and empirical connections to larger long-term goals. This is, in part, captured in the Netway and the ability it provides for cross-fertilization. In Phase II we will be continuing to examine these systems-level challenges.

1.2. Pilot Studies

The systems evaluation model and measures discussed here were developed, implemented and tested in 41 pilot projects at 9 sites that conducted formal and informal science education programs. The variety of settings included 7 county Cornell Cooperative Extension (CCE) associations, one large-scale NSF-funded STEM research and education outreach center, and one STEM museum. In the first year of the grant, we worked exclusively with Cornell Cooperative Extension of New York City to develop and test the Evaluation Planning Partnership (EPP). In Year two, six additional county extension offices in New York State were added as pilot sites and included St. Lawrence County (CCE St. Lawrence), Jefferson County (CCE Jefferson), Onondaga County (CCE Onondaga), Tompkins County (CCE Tompkins), Chenango County (CCE Chenango), and Ulster County (CCE Ulster). Two STEM-specific sites were also added in Year two: the Cornell Center for Material Research (CCMR; an NSF MRSEC site), and the Museum of the Earth/Paleontological Research Institute.


This method depends upon the creation and utilization of a partnership, which we termed “Evaluation Planning Partnership, or EPP. The focus of this work was simply on evaluation planning, and not on the implementation of the evaluation. An EPP consists of two groups of “experts”: the Evaluation Facilitator group [the P.I. and staff of the Cornell Office for Research on Evaluation (CORE)], and an organization. The organization is led by a director (frequently referred to as the Executive Director, or ED), an Evaluation Project Manager (EPM) - a member of the pilot site organization who served as the primary contact and liaison between the Evaluation Facilitators and the staff at their program site – and the program staff. The EPM was also responsible for collecting and submitting all supporting information necessary for the evaluation planning meetings; establishing a training calendar for all their pertinent staff; scheduling and chairing internal interim planning meetings as needed; scheduling, coordinating and hosting all interim teleconferencing and remote meetings and/or trainings pertinent to their site; and presenting their site specific results to other EPP pilot sites. Each Extension office selected one person to act as the EPM for their site. CCMR and Museum of the Earth – due to their small staff size - decided to take a more collaborative approach and had two to three staff members serving as the sites’ EPMs. By engaging in the Evaluation Planning (EPP) process, as of July 21, 2008 these sites had collectively generated 41 evaluation plans.

After our initial year working with NYC Extension offices we created an initial Systems Evaluation Protocol and accompanying educational resources, as well as a pre-EPP measure for assessing a site’s evaluation capacity. These assessments and protocol were used in Year 2 with 6 additional CCE sites, as well as the two STEM sites. During the second year we developed and implemented rubrics for assessing the quality of both the logic/pathway models and the evaluation plans of these 8 sites. Analysis was not
conducted on the NYC sites, although they continued to participate in our activities throughout the entire grant cycle. Post-EPP interviews with organizational directors, project managers, and staff were conducted by a non-affiliated Cornell laboratory (the Survey Research Institute, or SRI). These results will be discussed in greater detail in the section on project findings.

Finally, deeply integral to the EPP process has been the involvement of Undergraduate Research Assistants. Supplemental funding was received for supporting the REU experience. Students’ status ranged from freshman to senior, and their majors varied across the sciences, social sciences, and arts. The undergraduate research team members all expressed feeling overwhelmed by the scope and complexity of our project at first glance because this was their first real-life research experience. In their own words “the concepts of evaluation and measures seemed very abstract and elusive”. Over the course of the project most reported having gained a lot of valuable knowledge and experience, including teamwork with peers and supervisors, and knowledge about conducting evaluation and research. Before long they were able to work with minimal supervision and maximal collaboration. This project presented a wide range of challenges that required many different skill sets, so tasks were assigned according to students’ interests, including: web development, statistical analysis, literature searches, writing, collaboration, and management. They learned to coordinate with each other, as well as to collaborate and communicate with supervisors and managers.

Depending upon the stage of the project at the time of their involvement students gained varying levels of knowledge of the evaluation process: planning, measurement and analysis. One of the most interesting and rewarding aspects as reported by students of this project was being able to see real world applications of research (applied research), and their individual roles within that entire process. Most students worked part-time throughout the school year. One student worked one-on-one with a pilot site, which gave her a unique perspective of a developing program still putting together their evaluation plan, and her skills in helping to orchestrate change. Another student - from a non-research based college - was exposed to the university environment through a full-time summer position, and learned that she could contribute to conversations and research with more experienced researchers.

Students report that 1) they have gained invaluable skills in database research and utilization of other online resources; 2) they have learned to approach research in a systematic manner, including record keeping of the research process, 3) this experience contributed to fine-tuning of their writing, critical thinking and analytical skills, 4) it has broadened their perspectives on what is possible in research—not just setting in a laboratory environment—as well as what is possible for evaluation, and 5) that the while very challenging the research opportunities can also very rewarding and enriching.

1.3. Creation of Tools and Resources

Six key resources were developed as part of this project. These resources were based upon theoretical perspectives and provided the necessary mechanisms for implementing and facilitating the Evaluation Planning Partnerships. These resources were: (1) the Systems Evaluation Protocol (SEP) presented in the form of a Facilitator’s Guide; (2) The Netway; (3) the Cornell Office for Research on Evaluation (CORE) website; (4) the EPM network list-serve; (5) associated measures; and (6) program leader educational resources on evaluation planning. As mentioned in the previous section these will be discussed in greater detail in the section on products.
1.4. Theory Development

A major initiative of this research project was to develop cutting edge theories about evaluation and to apply them to evaluation practice. These theories have shaped the development of the products and resources discussed above.

1.4.1. Evolutionary Theory: Program and Evaluation Lifecycles.

The foundational lifecycle framework 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 human knowledge, including 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. Program variations are tried and survive or not according to socially negotiated selection mechanisms. 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.

1.4.2. Systems Theory:

While utilizing the SEP as a guide to evaluation planning results in evaluation plans that meet current best practices, the protocol was also explicitly constructed to extend current evaluation practice through deliberate integration of principles from systems theory. 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 consideration 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); nonexperimental, quasi-experimental and randomized experimental designs (Cook & Campbell, 1979); (Shadish, Cook, & Campbell, 2002); and so on. It is consistent with existing widely used evaluation models (Centers for Disease Control and Prevention, 1999) and with other approaches currently used in STEM evaluation (Frechtling, 2002).

The topic of systems thinking and systems theory is extremely complex; the literatures that informed the formulation of the SEP were considerable (refer to Phase II proposal for more information on this). The SEP integrated principles associated with these theories in order to assure that programs will incorporate such principles when: developing 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. During the integration of these theories into our project we made several presentations on the Theory of Systems Evaluation, and these are outlined in the “Presentations” section. Additionally, several pertinent publications are in the final stages of development.

1.5. STEM Evaluation Throughout the Life Course
An important collaboration within our research has been with Dr. Stephen Hamilton, a communication liaison between the SEP research group and two other groups: Cornell STEM outreach providers and New York State youth development practitioners. An early activity of his research was the solicitation of evaluation reports from Cornell STEM outreach programs. Most of these reports were simple compilations of data on audiences and activities with summaries of participant feedback based on end-of-program questionnaires.

He also organized sessions with STEM outreach providers in which their staff could identify needs and communicate project developments among themselves. These occurred on campus, January 17, 2006 and April 22, 2008. The first one alerted outreach providers about our project and created interest in both learning about and participating in the research. The second was a report on what the project had accomplished, featuring the pilot activities with the Cornell Center for Materials Research. STEM outreach providers found the report to be very promising as an approach they could use, not only for evaluation but also for program design and planning.

Evaluation is critical for making a link between research and outreach programs. In conjunction with the New York State Office of Children and Family Services, members of our research team helped organize and conduct two events on the topic of linking research and youth development programming. On September 25-26, 2006 both front-line practitioners and state agency representatives in Albany discussed issues inherent in the link between research and practice. A second event was held in conjunction with the New York State Youth Bureau Association on May 9, 2007, in Rome, NY.

A third event was a 1-day conference held at Cornell University for Youth Development professionals in Cooperative Extension. Here we presented the Systems Evaluation methodology to Cooperative Extension at large, and leading expert Dr. Nancy Trautman conducted two web-based surveys that focused on graduate student involvement in STEM outreach. Sarah Hertzog, a member of CORE, used the data from these surveys to establish the psychometric properties of the scales so that they can be used by other researchers. Ms. Hertzog has a paper in preparation that details this work. In addition, the Cornell Center for Materials Research (CCMR) is conducting secondary analyses on the Trautman data in order to evaluate the quality of STEM outreach for graduate students who have been involved with CCMR. Finally, we worked with the IGERT program to develop a concept map of graduate students’ views on the program as a whole.

We used a collaborative web-based methodology known as concept mapping to accomplish this assessment. Seventeen current IGERT students brainstormed 99 statements online in response to the following focus prompt: "One aspect or activity of the IGERT program that was valuable to me was...". Each participant then sorted the statements into piles of similar ones and rated them on a 1-to-5 relative importance scale. The sort data were analyzed with a sequence of multivariate analyses including multidimensional scaling and hierarchical cluster analysis (see Kane and Trochim, 2007 for details) and initial maps of results were prepared. These maps were interpreted by program participants in a facilitated session. The results described the major categories of program outcomes that were identified by the student participants and indicated which were relatively more important than others. The results were used in improving the IGERT program locally and were presented to NSF as part of their regular reporting. For purposes of our research, this pilot project illustrated that a systems-based evaluation
approach could be accomplished easily and quickly at low cost over the web for formative evaluation of graduate training in the sciences.

2. Findings

This project can be classified, according to its lifecycle phase, as a phase one exploratory study of an evaluation protocol and its corresponding cyber-infrastructure and resources. Due to its lifecycle phase, the feedback cycles were short and provided mechanisms for rapid adjustment and change. Thus, the findings reported here are primarily descriptive and relate to the development and implementation of this pilot test. Some of our findings relate to our experience in thinking about, developing, and implementing the Systems Evaluation Protocol (SEP). Other findings are more specifically linked to the degree to which the SEP and the Evaluation Planning Partnerships (EPP) enhance evaluation capacity in organizations that provide STEM education programs.

Several important and notable conclusions were reached during the development phase of this project. Based on our initial analysis and application of theory (including systems thinking, evolutionary theory, and developmental systems theory) we determined that a systems perspective has the potential to significantly enhance evaluation. Evaluation is an on-going process and should therefore occur throughout the life of the program and not only at the end of the program as is commonly done. When viewed as an aspect of program development, as opposed to a form of judgment, evaluation serves the function of providing constant feedback for program planning and improvement. Different evaluation methods are suitable depending upon the maturity of the program.

We also determined that the Evaluation Planning phase is likely to be the most critical step for STEM education evaluation because this is when the systems approach is most actively integrated into the standard processes of STEM programs. An investment of time and resources during the evaluation planning phase leads to high quality evaluations and findings.

We found that a formalized and standardized evaluation protocol is needed in order to develop consistent and high-quality evaluation plans. A standardized protocol allows the same evaluation planning process to be implemented across a wide range of STEM programs with a high degree of consistency and fidelity. Although facilitators tended to proceed through the steps of the protocol at various speeds and in various orders, all of the participating organizations arrived at the same end result. Additionally, the use of a standardized protocol addresses the inconsistent levels of expertise and experience that evaluators and practitioners may have with evaluation. The consultation/partnership approach used in the EPP process was effective. Using an evaluation expert and a program-level facilitator (EPM) enabled non-evaluation experts to take leadership roles within their own organizations regardless of their level of prior evaluation expertise. The facilitated process also assisted EPMs in acquiring a basic understanding of evaluation methodology.

The primary construct of interest for this project was the degree to which there is evidence suggesting an increase in evaluation capacity in partnering organizations. One of our greatest successes has been developing, implementing, and analyzing the results from measures and tools that we created specifically
to examine capacity building. These measures were applied across a wide range of STEM programs and delivered demonstrable evidence of the ability to successfully build organizational and programmatic evaluation capacity through careful and well thought out evaluation planning. These results were derived from pre-post analyses of organizational evaluation capacity.

2.1 Pilot Projects

Organization Capacity Survey:

We have worked with the evaluation of nine organizations, and the information summarized here was collected from six of them (five county Extension sites and one STEM museum).

The first category of questions on our organizational survey dealt with the basic structure of each organization. Two sites had staff trained in evaluation involved with their evaluation activities. Four had untrained staff on their evaluation team. Two sites utilized an external evaluator on a routine basis.

We then asked each participant about their evaluation budgets. Three sites reported an intermingling of their budget with their evaluation costs. The evaluation budgets ranged from $700 to $10,000 annually, but the sites varied enough in scope and size so more information would be needed in order to obtain a helpful understanding of any comparison.

In regards to training and technology, none of the organizations previously had in-service evaluation training. Every organization had a computer network offering both Microsoft Office and database software. Three organizations had specialized database staff. One had statistical software available, but none of the organizations reported having qualitative software at hand. Two sites had identified evaluation as a job requirement, but only one site reported having a requirement for written evaluation reports, and this site also has staff reviews based on evaluation data. None of the six organizations that responded had requirements for an evaluation-related budget.

Program survey:

We designed a program survey as a way of creating a baseline description of new programs that had not yet been involved with our SEP training. We received descriptive information from 28 programs from 6 county extension programs and 2 official STEM sites. These programs varied greatly.

Out of the twenty-eight programs, 23 had devised a formal description, and 22 had developed goal or mission statements prior to this research. Nine programs had less than a 10% FTE staff person, two more had staff FTE between 10% and 25%, four programs had a 26%-50% FTE staff person, and two had between than a 50%-99% FTE staff. One program had 150% FTE, one had 330% FTE, and one had 330% FTE. The remaining 8 programs declined answering that question.

One program reported a budget of over $150,000 and on the other end of the budget spectrum, three programs reported having $5,000 or less annually. Three programs had budgets between $5,000 and $19,000. Seven programs reported having an annual budget of $20,000 to $49,000. Five programs had a budget somewhere between $75,000 and $149,000.

Three programs catered to a population of twenty-five or fewer people, but the majority (18) served an audience of twenty-five to 500 people. Only three programs had more than 1,000 participants. Youth
Programs and Agricultural Programs were the two main areas of focus, but families and individuals with low socioeconomic status were also targeted groups. The typical age range for most of the programs was Children (ages 5-12) and Young Adults (ages 12-19). Sixteen programs were aimed at individuals older than nineteen, and only four programs had a target audience of five years old or less. The gender of participants seemed evenly proportioned, with 9 programs reporting a slightly higher number of females, 7 programs reporting a slightly higher number of males, and only 1 program reporting overwhelming disparity (more than 75% males). Most programs had white audiences (23 programs), but 13 programs reported serving African-Americans, 6 programs serving Latino audiences, and 4 programs – all in Ithaca, NY near the Cornell campus – also targeted Asians. Only two programs worked with individuals of American Indian or Alaskan descent.

We asked the program leaders if their programs reached beyond their home county or was strictly local. The response was overwhelmingly the latter. Seventy-five percent of the responding programs cater to only the local community, and only 4 programs (which included both of the STEM programs) were multi-county.

At the outset of this study 23 of the 28 programs were already being evaluated. Of these, 16 programs were being evaluated by an internal source, and 16 were being evaluated by external parties. (Ten programs used both methods). Half of all programs formally shared the results of their evaluations with others. We also asked participating programs about their chosen evaluation instruments, but only nineteen of the twenty-eight programs responded. Out of those nineteen programs, 9 used satisfaction surveys, 7 used pre- and/or post-tests, and 1 used interest surveys. In addition, 6 used another (unidentified) type of survey and 5 used other methods.

Twelve of these programs had begun to analyze their data. Non-program staff was helping four programs with their analysis, two programs were being analyzed by an educator, and two others by program managers. Eleven of the nineteen groups that responded had developed some sort of plan to report their data. One program reported obtaining approval of human subjects.

Lastly, we asked the programs about their sources of funding. Twenty-four of the twenty-eight programs responded to this set of questions. Of these twenty-four programs most relied on more than one source of funding. Fifteen rely on funding from their county and Cornell Cooperative Extension, 6 programs make use of state funding, 3 access city funds, and two utilize federal funds. Seven programs get support from donations and user fees.

**Rubrics**

In addition to our successful development of innovative tools to assess evaluation capacity, we also developed two rubrics that assess the quality of program logic models/pathway models and of the evaluation plans. Figure 1 shows the average ratings (out of 100 possible points) of all the logic models and evaluation plans within each site. Each of the county Extension offices also did their own internal ratings. We hesitate to speculate on interpretations at this point, but this measures offer a baseline for looking for improvement throughout the next phase of this project. Because the sites did not have official evaluation plans before this project it was not possible to take a true baseline.
EPP Summary
We held a panel discussion of the Evaluation Project Managers (EPMs) at our Youth Development Evaluation Conference in May 2008.

Evaluation Planning Partnership successes were identified by the EPMs included: a Partnership was a highly participatory method of evaluation which brought more people together (and thus more ideas); an increased ability to challenge the routine that many programs fall into; providing time for reflection; learning about other programs within the organization (thus creating community); and encouragement to form partnerships with other organizations.

There were also several challenges that were clarified by this study, such as: program staff not valuing evaluation or understanding its importance to the program; difficulty meeting deadlines or understanding the timeline process that was imposed by external forces; lack of understanding the time commitment; and fear of evaluation, as if evaluation would lead to a penalty of some sort.

Participants reported being surprised at the revelation that evaluation is an ongoing process, and not just an afterthought. Another surprise was the efficiency and usefulness of the Netway system. The Netway proved to be an easy-to-use tool that assisted many of the EPP leaders. Finally, a major surprise for a few of the panelists was how often they had to refer to past documentation while completing evaluation tasks.

The entire focus of this study was on evaluation planning. The panelists were asked why they thought so much emphasis was placed on planning, and if that emphasis was beneficial to the evaluation process. They said that the several planning stages emphasized thoughtfulness, encouraged enlightening discussion, and provided clarity.
**End-of-Project Survey**

In July of 2008, we contracted with the independent Cornell Survey Research Institute\(^1\) to conduct a telephone survey of the nine (9) pilot sites participating in Evaluation. One purpose of the survey was to assess the effectiveness of the EP and the CORE Evaluation Team in helping the pilot sites to build evaluation capacity. The following summary lists the number of respondents who mentioned each item, by job role.

<table>
<thead>
<tr>
<th>Challenges, negative effects and least useful aspects</th>
<th>SRI qs 1,3,5</th>
<th>(EDs not asked qs 3 and 5)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time demands</strong></td>
<td><strong>Structure didn’t fit needs</strong></td>
<td><strong>Data entry demands</strong></td>
</tr>
<tr>
<td>EPMs (^2) (N=8)</td>
<td>7+2+2</td>
<td>2</td>
</tr>
<tr>
<td>EDs (^3) (N=9)</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>Program Area Leaders (N=13)</td>
<td>7+7</td>
<td>4</td>
</tr>
<tr>
<td>Program Staff (N=11)</td>
<td>6+5</td>
<td>2</td>
</tr>
<tr>
<td><strong>TOTAL</strong> (N=41)</td>
<td>46</td>
<td>8</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Benefits, most useful effects, recommend for others</th>
<th>SRI qs 2,4,6</th>
<th>(EDs not asked qs 2 and 4)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Clearer thinking about program &amp; goals</strong></td>
<td><strong>Better able to articulate program to stakeholders</strong></td>
<td><strong>Better understanding of and appreciation for evaluation</strong></td>
</tr>
<tr>
<td>EPMs (N=8)</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>EDs (N=9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program Area Leaders (N=13)</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Program Staff (N=11)</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td><strong>TOTAL</strong> (N=41)</td>
<td>11</td>
<td>15</td>
</tr>
</tbody>
</table>
We are currently reviewing the full set of suggestions for improvement from these survey results and incorporating them into its 2008-2009 EP protocol, procedures and instruments. A summary follows, with the number of comments on each item in parentheses. If there is no number listed, the number is one.

<table>
<thead>
<tr>
<th>7. Suggestions for improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>SRI q 29, EDs q 47</td>
</tr>
<tr>
<td>Most mentioned</td>
</tr>
<tr>
<td>EPMs (N=8)</td>
</tr>
<tr>
<td>Executive Directors (N=9)</td>
</tr>
<tr>
<td>Program Area Leaders (N=13)</td>
</tr>
<tr>
<td>Program Staff (N=11)</td>
</tr>
</tbody>
</table>

(1) The Survey Research Institute is a full service survey research institute facility that conducts surveys and provides survey research services to Cornell University faculty, students, and administration, federal, state, and local government agencies, other nonprofit organizations, and other organizations. (http://sri.cornell.edu/sri/aboutus.vision.cfm)
(2) EPMs—Evaluation Program Managers
(3) EDs—Executive Directors

2.2 Netway

The Netway is the title of the web-based tool developed for use by Cornell Cooperative Extension, and significantly enhanced through the support of the National Science Foundation. As mentioned earlier, the user community was surprised (happily so) at the contribution of the Netway to their evaluation work, as well as the ease of use. The Netway adds significantly to the field of evaluation by assisting in the drawing of a pathways model of how a program works.

The Netway allowed us to easily compare the programs entered into it. For example, we were able to easily plot the lifecycle phase of all of the programs simultaneously which allowed us to essentially have a birds-eye view of the current status of STEM education programs in the Netway. As discussed above, we have developed a method for conceptualizing the maturity of a program based on the program’s stability as opposed to its chronological age. Interestingly, we found that very few of the programs have
passed the second stage of the program lifecycle. This finding may prove to be particularly important for future research.

Despite the clear positive contributions of the Netway, an unexpected result from the current study was that high quality evaluation plans and logic models can be generated by following the protocol without having access to the Netway. We observed a “domino effect” whereby participating pilot sites worked with colleagues both within their own organizations (who were not part of this study) and even beyond their own organizations to share the knowledge that they gained from being part of this study. For example, members of CCMR were approached by the Cornell Institute for Physics Teachers (CIPT) (an NSEC site) and were asked to assist with their evaluation. CIPT staff had witnessed the successful and transformative evaluation planning partnership process that CCMR had engaged in and they wanted to be included. After completing her own evaluation planning process, the EPM from CCMR successfully facilitated the implementation of the Systems Evaluation Protocol with staff at CIPT. This was done without the benefit of the Netway and demonstrated the effectiveness of the Systems Evaluation Protocol when used in isolation and without the support of a cyber-infrastructure.

Additional evidence of a “domino effect” occurred within the Cornell Cooperative Extension pilot sites. Within participating organizations, EPMs have been approached by non-pilot site program leaders who have asked the EPM to lead them through the Evaluation Planning Partnership process. This has resulted in several programs being added to the Netway that were not in the original target group. Based on informal feedback from pilot site participants, we learned that program leaders and administrators came to appreciate that evaluation is an on-going process while participating in this process. The participants also found that it is both difficult and undesirable to consider program implementation without simultaneously considering evaluation. By engaging in the logic/pathway modeling process, practitioners came to realize that by clearly articulating their program’s logic and linking their longer term outcomes to a research evidence base, they are able to address the questions and concerns of their funders while still measuring shorter-term outcomes which are of greater interest to them and their daily program improvement.

A substantial amount of data was also captured through the Netway usage statistics and serve to demonstrate the effectiveness, usability, and accessibility of an evaluation planning cyber-infrastructure. Ultimately, the results suggest that an evaluation cyber-infrastructure provides a much needed infrastructure that avoids the repetition of effort and generates much needed linkages amongst STEM practitioners, researchers, and evaluators.

As of October 2008 there were 192 users on the Netway, and Figure 2 shows that roughly 75 – 100 unique users visit the Netway every month, and that from May-July they averaged about 2 pages per visit. Activity was higher earlier in 2007 when the first version was initially released and used. The peak number of visitors occurred at start-up, as well as in October 2007.

Figure 2.
Figure 3 zooms in on the monthly visits graph and breaks it out by week so we could more closely compare activity with interactions (such as trainings – which have been identified on the graph) between the evaluators and the EPMs. Also showing is the bandwidth use, which will be a reflection of the amount of work done while on the Netway.
Here we can see the two peaks identified in the previous graph, as well as an odd peak in December. This third peak, not surprisingly, correlates with a significant year-end report to the organizational Directors, but this one-time brief flurry of activity didn’t register as a peak on the monthly report. The first peak at start-up is understandable as people try something new in preparation for the first training (logic modeling), and the second peak (in October) follows a training on measures.

### 2.3 STEM Evaluation through the Life Course:

Analysis of the STEM Outreach program reports indicated the magnitude of the need for improved evaluation. The best evaluation work was done by Professor Marianne Krasny and Dr. Nancy Trautmann who were responsible for the GK-12 program, Cornell Science Inquiry Partnership. Their evaluation designs were robust enough that they were able to publish their findings. The other evaluations he reviewed provided little guidance to the programs in which they were conducted and none to other programs.

Discussions around the gap between the needs for evaluation in youth development and both the current state of evaluation in the field and the resources available made it difficult for Dr. Hamilton to work with youth development practitioners on the topic of systems evaluation and evaluation systems. He tried to paint a broader picture and to lead audiences to understand the potential value of the approach developed in the project, namely making it possible for practitioners to find on the web both previous research related to specific youth development objectives and valid instruments for measuring the achievement of objectives.

The evaluation of youth development and of STEM outreach suffers from a severe lack of resources. Funders almost always require evaluations, but NSF is unusual in allocating funds for the purpose. Without adequate funding, practitioners take responsibility themselves for evaluation, almost always without benefit of the scientific background and training they need. Another scarce resource is a scientific research literature to draw on when designing and evaluating programs. Evidence-based practice is increasingly held up as the ideal, yet the amount of evidence available in these fields, as compared to medicine, is vanishingly small. As a result, we have inadequately trained people trying to do evaluation on a shoestring in the absence of a well-established set of instruments, concepts, and associations. The situation is self-reinforcing: low-quality evaluations are not published, so there is little progress toward building up the field.

### 3. Training and Development

Please refer to Outreach for direct information on the formal Training and Development we conducted.

Internally we have continually educated ourselves on the current research in evaluation theories and methods through reading academic journals and books. We have also attended conferences and workshops for the evaluation community.

Because the main focus of this project has been to train educators about the process of evaluation, training and development has been an inherent part of all of our activities. This is explained more indepth in our Activities (pilot projects) and in the Systems Evaluation Protocol, itself.
4. Outreach

We have primarily engaged in two forms of outreach, one which targets other evaluators and researchers and the other which targets educators and the community. Due particularly to our affiliation with Cooperative Extension, we have wholly embraced their mission to bring the university to the community. We have given presentations at numerous professional organizations (see above). We have also focused our outreach efforts around STEM outreach educators. For example, we spoke at the 2008 STEM Outreach Educators conference. Our partners at CCMR have taken the lead and are spearheading educational evaluation efforts amongst the MRSEC sites. They have presented their evaluation work to their colleagues at conferences and are leading the evaluation working group.

In addition, the P.I. of this project was contacted by the Dean of the School of Industrial and Labor Relations at Cornell University and has been asked to work with their educational outreach programs on evaluation. In 2008, our outreach and education focus has been on 4-H youth development and particularly their national Science, Engineering, and Technology (SET) priority area. On May 20, 2008, we hosted a conference for New York State 4-H Youth Development leaders and Cornell Cooperative Extension educators. The goal of this conference was to provide an update on the status of our current work, to promote future interest in the research project, and to emphasize the need for quality evaluations of youth development programming. Our guest speakers included Richard M. Lerner, Professor and Director of the Institute for Applied Research in Youth Development, Tufts University; William M. Trochim, Professor and Director of the Cornell Office for Research on Evaluation (CORE), Cornell University; Sarah Hertzog, Graduate Research Assistant, Department of Human Development, Cornell University; Barbara Schirmer, New York State Extension Leader, Cornell University, and Stephen Hamilton, Professor and Associate Provost for Outreach, Cornell University.

The response to the conference was very positive. According to the results of our post-conference survey, participants reported that our biggest successes were providing an environment for networking with other 4-H leaders and providing time to interact with and hear from experts in the fields of youth development and evaluation. Participants gave the conference an overall average rating of 4.3 out of 5 points.

We have also been using our website as a way of reaching out to the public. (http://evaluation.cce.cornell.edu/)

In addition to our work directly with educators, several presentations were made during this project that served to present the theory, results, tools, and resources of this work to evaluation professionals, STEM educators, administrators, and community members.

Presentations


Several presentations were made during this project that served to present the theory, results, tools, and resources of this work to the evaluation professionals, STEM educators, administrators, and community members. The following is a list of these presentations:


5. Publications

We are hard at work on several publications related to this project. The first paper presents the theory work done on this project and specifically outlines the major thinking behind what we have termed “evolutionary evaluation”. The second paper examines the ways in which systems evaluation can program planning can serve as a mechanism for integrating research with practice. The third paper explores the findings of the current project as a form of capacity building within organizations. The first two articles are planned for publication submission before the end of the year.

Also attached, please find our current version of the Systems Evaluation Protocol. We are planning a limited run printing of this publication for use with our Phase II funding, and are waiting for further development of the material before an official publication.

Additionally, our Co-PIs have their own articles.


Nancy Trautmann has recently submitted an article on her analysis of the survey data, entitled "Professional Development of STEM Graduate Students - Realities and Goals."

6. Products

6.1 Systems Evaluation Protocol (SEP)

A major initiative of this research project was the creation of what is turning out to be the Evaluation Planning section of our Systems Evaluation Protocol (SEP). This tool was specifically created for Evaluation Facilitators working with STEM education programs with a focus on increasing the evaluation capacity of an organization. However, the SEP has applications well beyond the field of STEM education. This protocol outlines the steps followed by the Evaluation Facilitators in the EPP process. When followed, this series of repeatable steps can lead to the creation of a project logic model, a project pathway model and an evaluation plan. Program leaders and staff learned evaluation skills that are applicable to all the program activities of their organization. Their paradigms of program evaluation and development were broadened to encompass the greater system within which a program is embedded. As a direct result of the intense interest in this protocol and its resources, additional funding was sought (and awarded) in order to assess the effectiveness of this protocol. Additionally, it is the intention of this investigator to complete the protocol to include evaluation implementation and utilization. Although the protocol to date addresses only the planning phase of evaluation, we will refer to the currently abbreviated form of the protocol as the Systems Evaluation Protocol (SEP).
Paradoxically, the SEP is a standardized protocol that enables any project to develop a uniquely tailored evaluation plan. In this sense it addresses well the STEM education environment - which needs a standardization of evaluation approaches - while recognizing the enormous varieties of contexts within which STEM programming occurs.

The evaluation planning phase of the SEP includes three broad stages, each of which involves several prescribed steps. (Once complete, the evaluation plan will guide the evaluation implementation).

**Stage 1: Preparation.** The planning stage of the SEP begins with a “Preparation” stage that is intended to acquaint the participants with systems evaluation and to establish baseline information regarding the target program(s)/organization. The preparation stage results in several outputs, including a Contact List, Memorandum of Understanding, Program Survey, and Organizational Evaluation Capacity Survey.

**Stage 2: Modeling.** The second stage is intended to further acquaint the participants with the SEP, enhance their knowledge of evaluation concepts, use systems approaches to analyze the programs, and begin developing program logic and pathway models. The Logic Model stage results in several outputs, including a Map of Stakeholders, Lifecycle Charts, a Program Logic Model, a Program Pathway Model, and a Collection of Evidence.

**Stage 3: Evaluation Plan Development.** The third stage, “Evaluation Plan Development” 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 Development stage results in several deliverables including a list of Evaluation Questions, a Collection of Measures, and a Systems Evaluation Plan.

As mentioned previously, the next two phases of evaluation – implementation and utilization – will be developed as the pilot sites move into those phases.

### 6.2 Netway: The Evaluation Cyberinfrastructure

The SEP also builds on principles from contemporary cyberinfrastructure research (National Science, 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 cyberinfrastructure system developed in our current research called the Netway. The Netway is a Web 2.0 application that is consistent with second-generation web-based communities and hosted services, such as social networking sites, wikis, and blogs (Wikipedia.com, 2007).

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.

The term “Netway” is derived from the phrase “networked pathway” and refers to the multidimensional 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 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. 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 (Asim, Essegaier, & Kohli, 2000); (Asim et al., 2000); (Burke, 2000). Each new project model adds to the online network of such models and can be accessed by others on the Netway. This helps ensure that even projects with no direct contact with one another are able to benefit from each others’ experiences (Marathe, 1999). The Netway was a central tool in our implementation of the SEP (Marathe, 1999) – meaning that we used it to create a logic model and pathways model of how the SEP “works”. Similarly, 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, a pathway model of how their program “works”, and an evaluation plan.

A future goal of the Netway is to allow funders (e.g., NSF) to view meta-summaries of projects across program areas, see where they are in their developmental lifecycles, and more effectively manage their portfolios of evaluations.

6.3 Website: Evaluation of Outreach and Extension
Originally we utilized our Extension website (http://evaluation.cce.cornell.edu/index.html) to provide general information on evaluation and specific resources for the network of staff, faculty and students at Cornell University and Cornell University Cooperative Extension who are consumers or providers of evaluation for extension or outreach programs. The website provides basic introductory materials on what evaluation is, when and why it would be used, how to accomplish it, and links to more detailed resources. However, the request for access to these materials is strong, and a new website is under development for dissemination of the results of this research. Stay tuned to the previously mentioned website for a link to the new website, currently planned for release in December 2008.

6.4 Network Development
Two major networks have resulted from this project. The first is the EPM list-serve (hosted on Google Groups) which, other than the Netway, has served as the primary communication mechanism across pilot sites. Access to this site is limited to the Evaluation Partnership leaders (EPMs and the Evaluation Facilitation Team). The EPM list-serve has 19 members and has encouraged collaboration between the EPMs. As of July 22, 2008 there had been 147 emails and 30 files posted to the group.

A second network, the Cornell Evaluation Network (CEN) was inspired by the current research and encompasses professionals, faculty, and students at Cornell University who have an interest in evaluation. This network came into being in November of 2006, and as of July 2008 consisted of 93 members and 116 messages. During this grant period the CEN held 12 campus-wide meetings, and plans are underway to continue these meetings in the 2008/2009 academic year. This networking group has its own list-serve that has primarily been used as a resource for connecting with evaluators who share common questions, as a mechanism for gathering feedback on presentations and reports, and as a means of advertising evaluation related opportunities such as conferences. (See http://evaluation.cce.cornell.edu/cen.html).

At the current time it is anticipated that a third network will be formed, with this one being for Cornell Cooperative Extension personnel. The EPMs who are participating in the present study are also expected to participate on this site and to disseminate their knowledge, skills, and resources to the NYS Extension community at large (see http://evaluation.cce.cornell.edu/een.html).
6.5 Associated Measures
Several measures were developed to assess evaluation capacity in the STEM system. These include: 1) An organizational capacity survey to assess evaluation capacity before implementation of the SEP; 2) A program survey to assess specific programmatic characteristics; 3) Rubrics for assessing the quality of logic models and evaluation plans; and 4) Phone Interview Questions for assessing change in organizational and program evaluation capacity.

The Program and Organizational Evaluation Capacity Surveys were designed to obtain baseline information and prepare an organization to become an evaluation partner. These particular surveys gather data on a program’s general information, participants, activities and desired outcomes of an evaluation.

The rubrics were specifically designed to provide constructive feedback on the models developed during this process by providing an assessment of model and plan quality that can be used to gauge progress in evaluation planning. They examine factors such as a program’s mission and description, as well as the measures, samples and design of the evaluation.

6.6 EPP Educational Resources
Training materials were developed to accompany the SEP: EPP Launch Presentation, Resource Slides, Tips on Logic Models, Tips on Measures, Tips on Evaluation Plans, Data Analysis Templates, and EPM Measures Guide.

7. Contributions within and outside of discipline
The project has contributed 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 involved STEM education and outreach specifically, the completed models and measures for evaluation systems are useful in many substantive areas that extend far beyond the STEM context. Thus, this project generated new knowledge in the evaluation of STEM education specifically and explored new concepts in the evaluation field generally. In addition, the model and measures that were developed will enhance the instrumentation, networks, and partnerships developed to support both STEM evaluation and systems of evaluation.

More specifically, the field of evaluation will benefit from the incorporation of systems thinking concepts into evaluation as well as concepts from evolution. For example, the notion of program lifecycle phases and methods matching demonstrate the strengths and weaknesses of a program based on maturation as opposed to chronological age.

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. The Systems Evaluation Protocol and its accompanying tools and resources will facilitate the promotion of evaluation to professionals in a variety of disciplines.

The findings and contributions of our research extend beyond the field of evaluation to include the fields of education, social science, and policy. Our project is specifically contributing to the field of STEM education by not only generating high quality evaluation plans that will lead to effective evaluations, but
also by enhancing the network of STEM educators and leaders. Our approach also provides a mechanism for program leaders and educators to demonstrate to their funders how their programs are having an impact on long-term outcomes, such as an increase in the number of scientists. Conversely, our project enhances the ability of funders and policy makers to see how their portfolio of projects converges on outcomes of interest. This ability has the potential to influence the funding priorities of agencies such as NSF. This work is also likely to trigger a closer examination of local and federal policies around evaluation.

This project also contributes more broadly to the social sciences through its examination of evidence-based practice, translational research, and effective research-practice integration.

Dr Hamilton has been bridging the worlds of academic scholarship and the practice of education and youth development throughout his career. In 1980 he published an article on the evaluation of experiential learning programs and one for practitioners on “Evaluating Your Own Program.” He brought this experience to this project, which has, in turn, deepened his understanding of evaluation systems and systems evaluation and yielded a new set of tools that he can share with practitioners, most notably the Netway. In his role as associate provost for outreach, he has helped to interest staff of several Cornell STEM outreach programs in the project and its products. He will continue to advocate for and facilitate both stronger evaluation and a firmer connection between outreach practice and research.

**Cornell Program Partnerships**

Through the various Cornell partnerships that have been created and nurtured throughout this research project, we have been able to address the need for opportunities in STEM education across the life course. These partnerships target children of all ages and adults. The Cornell Center for Material Research (CCMR) provides STEM outreach activities for children as young as five as well as opportunities for graduate students and faculty to further develop their educational outreach skills. By partnering with CCMR, the Museum of the Earth, and 4-H Youth Development through Cornell Cooperative Extension, we are able to help bring quality STEM education to the community.

We have also had an impact on STEM graduate education at Cornell University.

This project has directly resulted in a sustained network of STEM educators across New York State. This network of providers has built evaluation capacity which they are already sharing with colleagues who were not directly involved in this project. By establishing such strong connections, educators are able to exchange ideas and findings, thereby strengthening STEM education programs.

In addition, all of our pilot work in eight separate organizations, each of which has multiple programs, has involved training the staff in these organizations in evaluation and assisting them in implementing what they have learned.

Our contribution to human resource development has been indirect via our work with partnering organizations. Several of our pilot sites receive NSF funding and are therefore obligated to demonstrate their ability to provide opportunities for research and teaching in science and engineering, to demonstrate improvements in the performance, skills, and attitudes of members of underrepresented groups, and to develop and disseminate new educational materials. By partnering with us, these pilot sites are now able
to accurately and completely capture how their educational programs address these goals. Our approach to evaluation addressed the common challenges that NSF-funded STEM educational programs confront by helping educators to clearly conceptualize and articulate their program logic, link their local-level concerns with the more global level priorities of their funders, and to develop sound, scientifically validated methods of capturing and measuring their program’s ability to affect change.

Two graduate students have been involved with this research and have given several presentations at professional conferences as a result of this work. They are also in the process of writing several papers for publication. One of the graduate students is pursuing an academic career and plans on continuing to work in the area of STEM education evaluation.

Finally, several of our pilot sites have programs specifically aimed at groups that are underrepresented in STEM fields. For example, the Cornell Center for Materials Research (CCMR) has science outreach programs with schools in the Harlem Children's Zone and with the Tuskegee Institute. Some of these outreach programs have never been evaluated before. Through our partnership with CCMR and with Cornell Cooperative Extension offices that have similar programs, we are developing capacity within those organizations to evaluate their efficacy of their programs and to identify areas or ways in which they might better serve the populations with which they work.

The web-based Netway evaluation planning and tracking software that we have developed will be available in the future for either commercial or public use, and would be applicable across the entire spectrum of STEM education and outreach programs.

And more broadly, work on this project has spilled over to our other activities and has led to the PI, who is the current President of the American Evaluation Association, to develop the 2008 AEA Presidential theme: Evaluation Policy & Evaluation Practice. The concept of Evaluation Policy also contributes directly to other projects of the PI, including several NIH funded projects.
References


http://jay.marathe.net/profile/articles/1999-11community.htm


www.margen.com/6Ymas/Documentos/640108_community.htm


http://en.wikipedia.org/wiki/Collaborative_filtering