AC 2007-2304: THE EVALUATION OF A COMPREHENSIVE MIDDLE SCHOOL OUTREACH PROGRAM--THE STRATEGY, THE RESULTS, AND THE CHALLENGES

John McLaughlin, MacLaughlin Associates

John McLaughlin is a senior consultant in strategic planning, performance measurement, and program evaluation. He is presently working on several project including the Environmental Protection Agency, Health and Human Services Administration on Aging and Centers for Medicare and Medicaid Services, as well serving as the lead evaluation consultant to seven national centers.

Gail Hardinge, College of William and Mary

Gail Hardinge is an educational psychologist who currently works with the Va. Department of Education's Training and Technical Assistance Centers, at the College of William and Mary, providing professional development programs for teachers. She has worked in public education for twenty-two years and is an adjunct Assistant Professor at William and Mary, teaching courses in collaborative consultation and assessment, as well serving as the college's VDP Project Coordinator.

Eugene Brown, Virginia Tech

Eugene Brown is Professor of Mechanical Engineering at Virginia Tech. He has worked with ONR since 2001 on university-centered Navy work force development issues. He teaches undergraduate and graduate courses in thermodynamics and fluid mechanics and is the author of many papers and reports describing his research in the areas of computational fluid dynamics and aircraft propulsion.

Kirk Jenne, Office of Naval Research

Kirk Jenne is the director of the N-STAR (Naval Research--Science and Technology for America's Readiness) program at the Office of Naval Research. His on a rotational assignment from the Naval Undersea Warfare Center in Newport, Rhode Island. His research interests are in ocean engineering, materials, and underwater acoustics and sensors.

Robert Stiegler, Naval Surface Warfare Center, Dahlgren Division

Robert Stiegler is currently supporting the USMC Targeting and Engagement Systems and the Office of Naval Research, N-STAR initiative. His recent career experience has included service as a program manager for USMC science and technology programs, Science Advisor to the Commanding General, Fleet Marine Forces Atlantic, and Head, NAVSEA Combat Systems Safety and Engineering Division.

Evaluating a Collaborative Middle School Outreach Program--
The Strategy, the Results, and the Challenges

Abstract

A Navy funded outreach program directed at increasing the interest of middle school students in pursuing careers in science and engineering has been subjected to a comprehensive mixed mode evaluation process conducted by a professional educational program evaluator. The evaluation process and its implementation in terms of the comparison-group studies and mixed method (quantitative and qualitative) data collection methods are described as well as the logic models which underlie this process. Some results are described and the paper closes with a statement of the challenges faced by the evaluator.

The Program

In the fall of 2003, representatives from the Office of Naval Research (ONR) and the School of Education of the College of William and Mary met with the staff of Senator John Warner (R-VA) to seek funding for an innovative program devoted to increasing the interest of middle school students in pursing careers in science and engineering. This was in response to an acknowledgement by ONR that attention needed to be given to the future technical workforce needs of the Navy's Warfare Centers given the decrease in the numbers of US citizens obtaining advanced degrees in science and engineering[1].

From these discussions, the ONR-funded Virginia Demonstration Project (VDP) emerged which, in the three years of its existence, has grown to reach more than 1500 7^{th} and 8^{th} graders in its academic year and summer camp programs, to involve more than 80 science and math teachers in its professional development activities, and to employ the services of nearly 50 young Navy scientists and engineers who work side-by-side the teachers in the classroom as facilitators, mentors, and role models.

Figure 1. VDP Summer Camp, Dahlgren, Virginia, 2005

There were two types of problems used in the VDP: 1) Robotics Challenges requiring the solution of problems having both a societal and Navy focus using LEGO Mindstorms equipment and 2) specially-constructed Breakout Experiments which were used to illustrate specific mathematics and science topics.

The Robotics Challenges used the LEGO Mindstorms kits and the ROBOLAB programming environment and were modeled after the activities of the FIRST LEGO League (FLL), a highly successful program which operates robotics competitions both throughout the US and internationally. In the first year of the project, it was decided to develop a challenge around using robotics in a mine-clearing, rescue, retrieval, and construction operations. The societal connection was provided by the humanitarian aspects of the (then current) news of the tragedy of the tsunami in Southeast Asia which required the location and retrieval of land and sea mines that were scattered by the event. The establishment of a base of operations and the rescue of a damaged submarine and stranded persons at sea were included in the overall challenge. In the second year of the project, seeking a project that was more life-science oriented, we developed the Grounded Oil Tanker Robotics Challenge that involved the containment of marine animal injury resulting from the oil spill from a grounded freighter. This scenario was again inspired by a natural disaster--the recent hurricanes in the Gulf region of the US.

The Breakout Experiments used in the summer-camp program included the Ball Drop, Paintball (which included high-speed photography), Water Balloon Launch, Electric Gun, Rocket Construction/Launch, Egg Drop, and Alarm System experiments. These experiments where selected based on our desire to provide a selection of projects that would demonstrate a wide variety of science- and physics-based phenomena, involve appropriate math and problem solving tasks, provide team building opportunities, and be of interest and fun to a diverse group of students both with regard to gender and ability. An additional (and important aspect) was that the experiments were to exhibit a close connection to the Virginia Standards of Learning (SOL).

It is has been implemented in the school systems surrounding the Naval Surface Warfare Center (NSWC) in Dahlgren, Virginia, and includes the sixteen middle schools in the Stafford, Spotsylvania, and King George public school systems and includes both in-school and summer camp programs that use problem-based learning to stimulate interest in science and engineering among 7th and 8th grade students. The School of Education of the College of William and Mary provides professional development workshops and instructional materials for the teachers and mentors training involved in the program. The Navy mentors from NSWC, Dahlgren, play a major role in training the teachers in the use of the LEGO Mindstorms kits and ROBOLAB software which is the focus in much of the classroom and summer camp activities.

The program is structured to contain the features which are indicative of best practices in K-12 outreach activities[2], and encouraging evidence has been gathered to indicate that the program is making a difference in the way students view careers in science and engineering. Likewise, many of the teachers involved in the program have accepted problem-based learning as their preferred method of instruction. Nevertheless, there was limited documented evidence of success in terms of how effective the project was in stimulating a greater long-term interest among middle school students in science and mathematics.

Midway through the project it became evident that we needed to examine what the core elements of the program were, what the nature of the “intervention” was, how we wanted to define success, and how we could collect information which would not only tell us how we could re-direct the program, but also, if necessary, how to increase its impact. With this in mind, the College of William and Mary contracted for the services of the first author as the program evaluator.

Program Evaluation

Consistent with the ABET's defining elements evaluation[3],

> <u>Evaluation</u> – *Evaluation is one or more processes for interpreting the data and evidence accumulated through assessment practices. Evaluation determines the extent to which program outcomes or program educational objectives are being achieved, and results in decisions and actions to improve the program.*

the process should consist of several critical components, including: identification of results, examination of relationships between project outcomes and process, as well as examination of how the context might have influenced the results. An evaluation should provide information regarding the extent to which the NSTAR VDP's objectives were achieved, as well as actions taken to improve the project. The VDP evaluation team's goal is to help the project managers and staff not only to collect reliable, valid, and objective evaluation information, but also to create understandings of why things happened the way they did and what might be done to improve the results.

The proposed evaluation is an inclusive, mixed methods design based on the principles of the transformative paradigm of research[4]. The role of the evaluator is to raise questions surrounding mechanisms for meaningfully involving diverse groups in the process of constructing and understanding of what the project should do and actually does, and documenting its accomplishments. A second underlying philosophical assumption is the need for an interactive relationship between the evaluators and the stakeholders for deeper understandings of diverse perspectives to emerge. Therefore the evaluation design includes mixed methods, with both qualitative and quantitative data collection methods[4].

Wholey[5] identifies evaluability assessment as one of the first steps in the evaluation planning process, noting that evaluations are likely to succeed in producing evidence that can be used for understanding and improving program performance if the following conditions are present:

1. Evaluators and intended users agree on the goals, objectives, side effects, and performance criteria
2. Program goals and objectives are found to be realistic given the resources available
3. Relevant information on program performance is available and accessible
4. Administrators on the policy or operating level are willing to change the program on the basis of evaluation information.

In an effort to increase the probability of the formation, implementation, and use of an effective evaluation, the evaluators addressed each of these factors with the program's management. After acquiring management commitment to supporting and using evaluation for program improvement and accountability, the evaluators turned to the first characteristic of evaluability by engaging staff in the process of logic modeling[6].

It might be said that it is an evaluation axiom that you should not evaluate a program that you do not know. In our opinion (there are other approaches, see Stufflebeam[7]) a program evaluation that is not based on an understanding of what the program is trying to do, for whom, and using what strategies, might be misguided and might result in an invalid portrayal of the program's performance[4]. An important step in evaluation is to make explicit the theory of change using a technique for describing the program such as logic modeling[6]. A logic model presents a plausible and sensible model of how the program will work under certain conditions to solve identified problems[8]. Patton[9] refers to a program description such as this as an "espoused theory of action", that is, stakeholder perceptions of how the program will work. Evaluators are often called upon to collect information that attests to the validity of the espoused theory of change. Does previous research support the theory of change? Is it reasonable to expect the proposed outcomes to occur given the activities described?

The elements of a logic model are resources, activities, outputs, short-, and intermediate- and longer-term outcomes. Some add the customers reached, as well as the relevant external contextual (antecedent and mediating) influences on program implementation and success. A simple logic model appears in Figure 2:

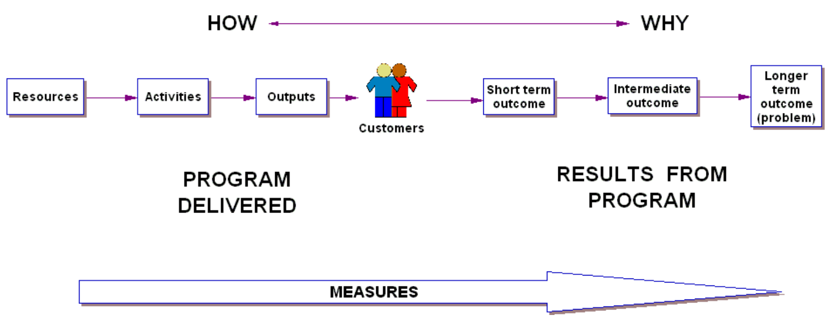

Figure 2. Simple Logic Model

The elements of the logic model are described below:

<u>Program Structure</u>

- *Resources/Inputs:* Programmatic investments available to support the program.
- *Objectives/Activities:* Things you do– activities you plan to conduct in your program.
- *Outputs:* Product or service delivery/implementation targets you aim to produce.
- *Customer*: User of the products/services. Target audience the program is designed to reach.

<u>Outcome Structure</u>

- *Changes or benefits:* resulting from activities and outputs.
  - Short-term (K, S, A) – Changes in learning, knowledge, attitude, skills, understanding
  - Intermediate (Behavior) – Changes in behavior, practice or decisions
  - Long-term (Condition) – Changes in condition

The evaluation design is formulated with the program's underlying logic structure serving as the guiding framework for the design and implementation of data collection strategies, data analyses, reporting and utilization of findings.

Developing the program's logic model is a social process in which the evaluator's role is to use interactive dialogue with staff to make explicit their implicit beliefs about how their program is supposed to work to achieve its intended outcomes. It is important throughout this process to consider the program's implementation context as it might influence the design, delivery, and success of the project. In the end, the evaluators will develop evaluation questions and performance measures for key elements along the logic model (from inputs through activities and outputs to outcomes) and develop data collection strategies to create evidence to answer the questions.

Evaluation seeks to establish understanding of performance from two perspectives. Best practice indicates that, after observing a particular level of performance (e.g., observed outcome levels for students participating in the intervention), the evaluator will retrospectively examine antecedent performance (levels of implementation of key intervention variables) to assess possible causal influences and looks at context to determine the degree to which variability in the outcome could be caused by some contextual influence. Then the evaluator engages the staff in prospective evaluation trying to understand the possible influences of observed levels of performance on subsequent performance. For example, if students' acquisition of certain knowledge variables is at 80%, what effect would this level of performance have on future behavioral applications?

The Logic Models

After reading background documentation on VDP (proposals, articles, previous evaluations) and conducting interviews with management and staff, the evaluator involved staff in the development of their program logic through a series of staff meetings. Three 'levels' of logic were unpacked started with a high level picture of the program in perspective to its resource pool and strategic impact. Next a level II model was produced that displayed the elements of the program in terms of these endpoints. Finally, a level III model was developed that showed how the program was supposed to work and key performance measures associated with this logic. The three levels of logic are presented in Figures 3, 4, and 5.

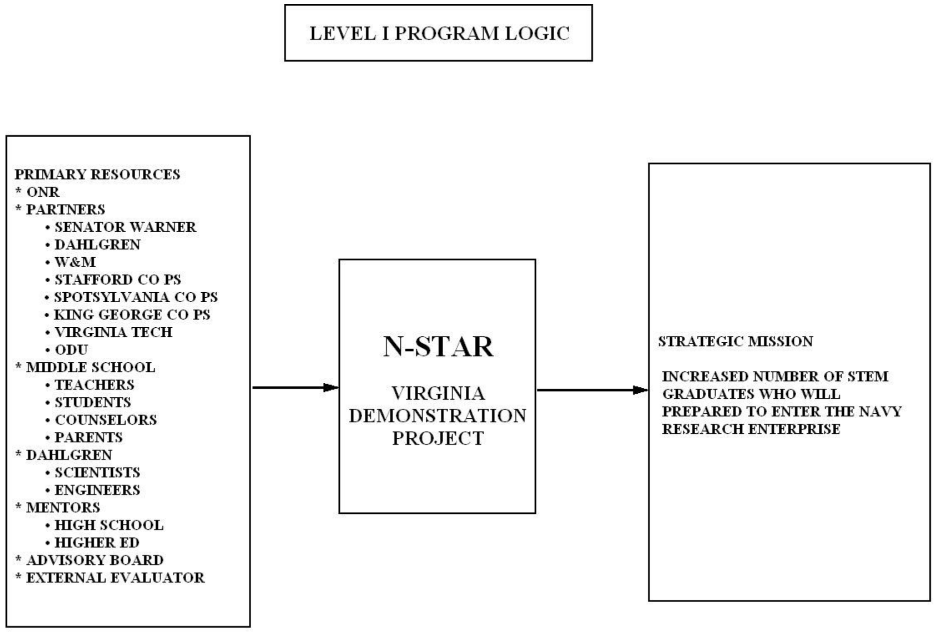

Figure 3. Level I Logic Model

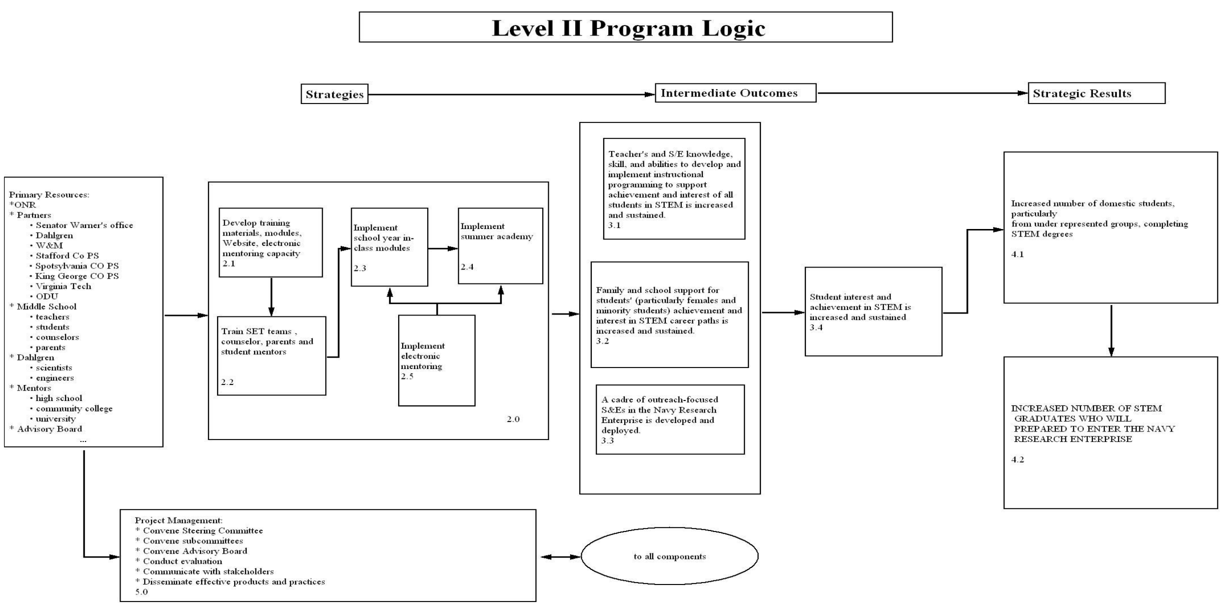

Figure 4. Level II Logic Model

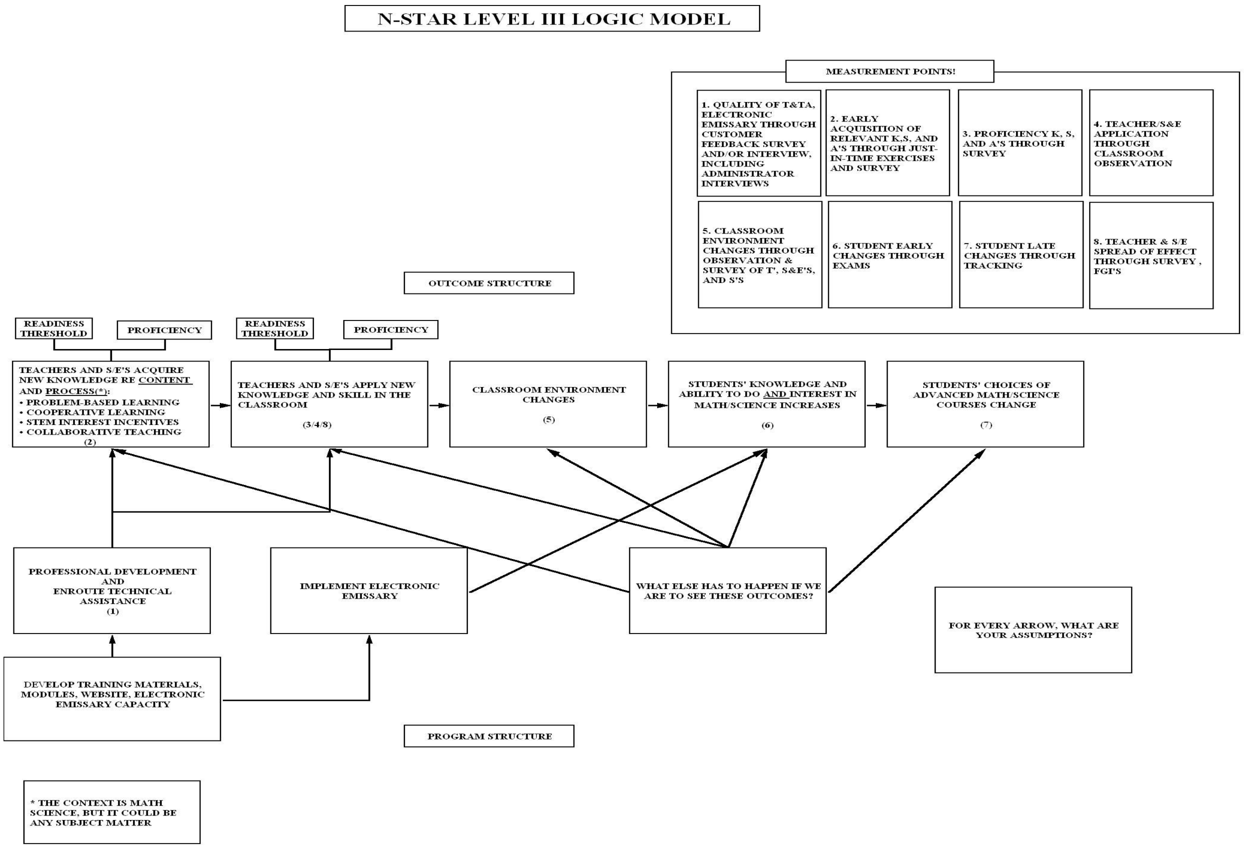

Figure 5. Level III Logic Model

N-STAR Evaluation Questions

In consultation with project staff the evaluator created the following set of evaluation questions:

1. Do the Steering Committee members sustain a high level of satisfaction with:

   - Shared vision for the VDP
   - Consultative decision-making
   - Shared leadership within and across levels
   - Communication: internal and external
   - Project management across levels
   - Ongoing performance measurement and evaluation
   - Future funding efforts for sustainability and diffusion

2. Implementation Fidelity

- Quality of professional development, technical assistance, and Electronic Emissary as measured through post-implementation feedback surveys and interviews
- Quality of classroom delivery as measured by classroom observations, interviews, and surveys administered during and post-implementation.

3. Outcomes for Teachers and Navy Mentors

- Early and on-going skills and knowledge change as measured by post-professional development surveys and interviews and observations, including embedded assessments throughout training.
- Changes in classroom behaviors including working with other professionals and students with respect to problem based and collaborative as measured by post-professional development surveys and interviews and observations.
- Spread of effect to other application settings measured by post-professional development surveys and interviews.

4. Other Outcomes

- Changes in counselor and parent knowledge, skill, and attitude RE guiding students toward STEM careers including the selection of appropriate courses as measured by post-intervention surveys and interviews.
- Changes in school climate and practices as measured by post-intervention surveys and interviews conducted with administrators and school staff.
- Changes in Community awareness and valuing of the VDP as measured by post-intervention surveys and interviews (individual and FGI).

In addition, the evaluator and staff recognized the need to develop opportunities for comparisons as follows:

- Participants (e.g., teachers, mentors, students) to themselves across time
- Participant performance to others who have not been exposed to the VDP
- Participant performance to participants in other similar programs
- Within and across school participants/non-participants talking into account context.

Data Collection

Through the spring of 2007 data will be collected in the three schools participating in the VDP. As noted above, a mixed methods evaluation approach will be utilized. Both qualitative and quantitative information will be collected extant data sources (school reports) as well as emergent data through observations, individual and focus group interviews and the administration of pre- and post-surveys focusing on satisfaction and knowledge and skill acquisition as well as behavioral changes. In addition, because of contextual variations across school divisions included in the program, an embedded case study approach will be employed where individual divisions served as nested cases in the three-school division program.

Evaluation Results

As the program unfolded (2004-2006), formative evaluation strategies were applied to create information for continuous program improvement in which the program staff and Steering Committee were the primary audience. In the final year of the program (2007) summative evaluation procedures were applied in which external audiences were the evaluation customers. The aim was to address the questions of worth as well as the extent to which the project should be disseminated to other settings with similar challenges.

By the end of the 2006 school year, 1678 students, 86 teachers and 48 scientists and engineers had participated in the VDP. Fifty school counselors had participated in the counselor education programs. Ten different sets of professional development training activities were provided to professional co-teaching teams.

The initial phase of the VDP yielded strength in program development and implementation. Six out of the seven initial project goals were met by the 2004-2006 VDP activities. Problem based learning modules were developed and delivered by teams of co-teachers consisting scientists, engineers and middle school teachers. Teacher reports indicated that professional staff gained knowledge, skills and abilities for supporting student interest in STEM careers. Students also reported an increase in interest in and enthusiasm for STEM careers. In addition to career information, students gained knowledge about science topics (landmines or coral reefs) and developed skill in programming, use of digital cameras and video editing software.

Eighty percent of scientists and engineers agreed that the VDP exposed students to science and technology, and that the experience could steer students toward STEM careers. Likewise, teachers reported positive outcomes, indicating that students were "active, enthusiastic, and engaged in the program's activities." Students surveyed revealed an increase in knowledge about robotics, an increase in knowledge about what scientists and engineers do in their jobs, and an increase in interest in STEM careers.

Student focus groups were held in four schools. Comments made by students reflected a high degree of satisfaction with the content and the opportunity to interact with scientists and engineers. Specifically, students noted the benefits of a problem solving approach, saying [It] "shows why you're learning stuff", and "you're having so much fun that sometimes you forget that you're learning." One student offered to the group that the VDP activities would be "good to have all over the country…it could change the way people think about things…like pollution problems and ways to stop it and fix environmental problems."

During the spring of 2007, the VDP will be in the final phase of implementation. Both professional development content as well as the curriculum modules will have been revamped. The curriculum is more defined and explicitly linked to Virginia's Standards of Learning. The professional development workshops are streamlined and have included suggestions made by teachers over the last twenty-four months. Specifically, additional time has been allotted for teacher planning and robotics training. Despite the decrease in S&E support, the teachers determined early in the school year that they wanted to proceed with the project. Interestingly, the teachers who had previously participated in the VDP took on a master teacher role, and offered to lead their respective schools through the project. Still, there was concern that an anticipated decrease in availability of Navy mentors due to an unanticipated shortfall in Congressional funding would clearly dilute the much-needed career emphasis. In the final version of the paper additional details about this part of the work will be given.

Evaluation Challenges

There are always challenges to the successful design, implementation and utilization of evaluations of any program. Several of these were discussed above under the topic of evaluability assessment. Additionally, evaluators need an explicit theory of change, often described using the logic model, good implementation fidelity of that theory, high quality measurement, and access to necessary data in a timely manner. The following challenges have been encountered in the evaluation of the N-STAR program.

1. <u>Shifting funding</u> - Over the course of the program, staff often had limited knowledge of actual level of funding that would be available to support both the program intervention efforts and the evaluation. In the final stage of the project there has been a significant reduction in funding that has a significant impact of the level of intervention available to staff.
2. <u>Political climate</u> – The VDP funding was obtained through an 'ear-mark'. The political climate at the end of the current funding cycle did not support the continuation of 'ear-marks' and as such limited the implementation of the final year's change strategies. The apparent lack of continuation prospects took the sails out of the evaluation effort.
3. <u>Management context</u> – The management of the program crossed three very different environments – the Navy, higher education and K-12 public schools. Each of these environments had its own expectations for project management, delivery, and success. While the public schools and William and Mary were anticipating a more linear approach to the design and implementation of the project, the Navy was accustomed to spiral development theory, which supported early implementation with changes made throughout the project.
4. <u>Intervention implementation limitations</u> – Due to contextual influences, both positive and negative, the intervention continually changed, often seeming to be something akin to 'building the airplane while it was flying'. Further, the incubation time for change varied and, in some cases was too short to expect change – the intervention exposed teachers to brief professional development and students sometimes experienced the intervention for only one day per week. Research on professional development suggests that in addition to an initial comprehensive training session, reinforcement of skills throughout the project is recommended in order to establish long-term change. Lastly, one of the primary goals of the program was to encourage students to consider STEM careers. However, the focus was, at times, implicit, and the reduction in mentor support further impacted the career focus of the project.
5. <u>Evaluation Instruments</u> – Despite considerable focus on attitudes associated with math and, to a lesser degree, science, a review of the literature has revealed only a limited number of

assessment instruments--Kristjanson[10], Walberg[11], Fouad[12], and the Bayer Facts of Science Survey[13]. In addition, the Fenneman-Sherman Attitudes Scales[14] has been used in a number of projects to evaluate male and female attitudes towards mathematics. This instrument is being considered as a measure for the spring 2007 VDP participants in addition to the College of William and Mary Observation Scales[15] which is currently being used. In addition, students will complete a pre/post Student Content Questionnaire, a Satisfaction/Efficacy Survey, as well as participate in mid semester and final Focus Groups. The professional staff will also complete a Content Questionnaire, a Teacher Efficacy Belief Scale and a Satisfaction Survey, as well as participate in Focus and Debriefing Groups. Finally, both counselors and administrators will participate in Focus Groups to determine the extent to which leadership and guidance have impacted the degree of success noted in the various school systems implementing the VDP.

Conclusions and Recommendations

The evaluation of the VDP indicates that it offers promising programmatic considerations through teacher training, curriculum development, mentoring, and staff support for program change. Teachers and school personnel have the opportunity to partner with scientist/engineer mentors to create an enhanced opportunity for career awareness for school staff, students and their families. Educators committed to bringing a problem-based, mentor-focused curriculum can build pathways to STEM career awareness by utilizing the lessons learned through the VDP. Specifically, it is recommended that school leaders can support grant initiatives to enhance data-base decisions for increasing mathematics, technology, engineering and science interests through programs similar to the one described here.

Bibliographic Information

(1) Brown, E.F., R.A. Kavetsky, R.L. Stiegler, P.N. Squire, J.J. Matkins, G.B. Hardinge, J.A. McLaughlin, "VDP--A Mentor-Focused Middle School Outreach Program," Paper 2006-991, ASEE Annual Meeting, Chicago, IL, June 18-21, 2006, http://www.asee.org/acPapers/code/getPaper.cfm?paperID=11587.
(2) Douglas, Josh, Eric Iversen, and Chitra Kalyandurg (2004). "Engineering in the K-12 Classroom—An Analysis of Current Practices & Guidelines for the Future," ASEE, November, 2004, http://www.engineeringk12.org/Engineering_in_the_K-12_Classroom.pdf.
(3) ABET, Engineering Accreditation Commission, "Criteria for Accrediting Engineering Programs, Effective for Evaluations During the 2006-2007 Accreditation Cycle," Approved October 29, 2005.
(4) Mertens, D., and J. A. McLaughlin, *Research and Evaluation Methods in Special Education*. Newbury Park, CA: Corwin Press, 2003.
(5) Wholey, J., "Evaluability Assessment." Chapter in Wholey, et. al., *Handbook of Practical Evaluation*, Jossey-Bass, 2004.
(6) McLaughlin, J. A. and G. B. Jordan, "Logic Models: A Tool for Describing Program Theory and Performance," Chapter in Wholey, et. al., *Handbook of Practical Evaluation*, Jossey-Bass, 2004.
(7) Stufflebeam, D. "Evaluation Models. New Directions for Program Evaluation," no. 89. San Francisco, Jossey-Bass, 2001.
(8) Bickman, L., "The Functions of Program Theory," In L. Bickman (ed.), *Using Program Theory in Evaluation*. New Directions for Program Evaluation, no. 33. San Francisco: Jossey-Bass, 1987.
(9) Patton, M.Q., *Utilization-Focused Evaluation: The New Century Text.* Thousand Oaks: Sage, 221-223, 1997.

(10) Kristjanson, Cheryl R., "Voices from the other side of the room: A study on changing teacher strategies to include girls in mathematics, science and technology," Proquest Dissertations and Theses, Section 0303, Part 0727228 pages, University of Manitoba (Canada); Publication Number: AATNN16182, 1996.
(11) Walberg, H.J., B.J Fraser, and W.W Welch, "A test of a model of educational productivity among senior high school students, *Journal of Educational Research*, 79, pp 133-139, 1986.
(12) Fouad, Nadya A., "Career Linking: An Intervention to Promote Mathematics and Science Career Awareness." *Journal of Counseling and Development*, May/June, Volume 73, 527-534, 1995.
(13) "Parents of Under-Represented Students in Science and Engineering Speak Out on Issues in New National Survey," *Bayer Facts of Science Education Survey,* May 24, 2005.
(14) Fenneman, Elizabeth and Shermann, Julia., "Fennema-Sherman, "Mathematics Attitudes Scales: Instruments Designed to Measure Attitudes toward the Learning of Mathematics by Females and Males," *Journal for Research in Mathematics Education*, Volume 7, No. 5, pp. 324-326, 1976.
(15) Van-Tasselbaska, Joyce et al., "Classroom Observation Scale-Adapted," The College of William and Mary, adapted from the Classroom Observation Scale-Revised," Center for Gifted Education Center, 2007.