I. INTRODUCTION

As part of its commitment to Congress to evaluate each of its programs periodically, the National Science Foundation's Directorate for Education and Human Resources (EHR) contracted with SRI for an independent evaluation of the Undergraduate Faculty Enhancement (UFE) program. This is the final report from that evaluation.

In addition to reporting descriptive information, this study provides the first systematic examination of associations between various characteristics of the workshops and their impact on faculty participants' subsequent development and implementation of courses.

Overview of the UFE Program

From its inception in 1950, NSF has supported various forms of professional development for faculty who teach undergraduates. The scope and variety of development offerings have changed over the years. By FY 1981, when funding for most of NSF's education programs was terminated by Congress at the President's request, three programs were providing direct, explicit support for faculty development. One of them, College Faculty Conferences, sponsored summer workshops over 3- to 4-week periods.[1]

During the mid-1980s, the National Science Board, Congress, and the White House all worked toward the reestablishment of NSF's education programs. Guided by a National Science Board report (1986), NSF created several new education programs and recast some former ones. Among the new programs was UFE, which was run by NSF's Division on Undergraduate Education (DUE).

UFE's goals were to support projects that enable faculty members who teach undergraduates to adapt and introduce new content into courses and laboratories, to investigate innovative teaching methods, to synthesize knowledge that cuts across disciplines, to learn new experimental techniques and evaluate their suitability for instructional use, and to interact intensively with experts in the field and with colleagues who are active scientists and teachers (National Science Foundation, 1997). UFE kept

[1] The other two were a sabbatical leave type of program and short courses of classroom instruction.

the concept of faculty enhancement through workshops from College Faculty Conferences, but shortened the time devoted to actual workshops and placed more emphasis on participants' preparation and follow-through. Projects could be regional or national in scope and could be in any field of science, mathematics, engineering, and/or technology (SMET). Many workshops used materials previously developed by Principal Investigators (PIs) under NSF Course and Curriculum Development grants.

The program operated from 1988 through 1998,[2] funding more than 1,000 workshops in all fields of SMET. In 1998, it was succeeded by a "track" within a larger program of course and curriculum development and laboratory improvement.[3] The new effort, called the National Dissemination track, focuses on disseminating exemplary course and curriculum materials and practices by providing faculty with professional development activities on a national scale. Workshops and short courses remain the primary mechanisms; the possibility of proposing to conduct distance-learning activities was explicitly added. Among NSF's stated expectations for the new track are to introduce new content into undergraduate courses and laboratories, to enable faculty members to explore effective educational practices, and to include participation by faculty who are representative of the national demographic and institutional diversity within the included field(s).

What Is Known about Workshops as a Strategy to Improve Undergraduate Education

Virtually all surveys regarding improving undergraduate teaching and learning have identified workshops as a primary strategy for this purpose. These include surveys of colleges and universities in the United States (Centra, 1978a, 1978b; Erikson, 1986; Hellyer and Boschmann, 1993), some of which focused on community colleges (Hansen, 1983; Richardson, 1987; Smith, 1981), and an international survey involving respondents in the United States, Canada, the United Kingdom, and Australasia (Wright and O'Neil, 1995).

[2] A small number of workshops were held in 1999.

[3] The Course, Curriculum, and Laboratory Improvement Program, Track 3: National Dissemination (CCLI-ND), described in NSF Program Solicitation 00-63 (May 1, 2000).

Despite the widespread use of faculty development workshops, there is relatively little empirical evidence regarding their impact other than in anecdotal form. Eison and Stevens (1995) observed:

> [C]ountless faculty development workshops are conducted on college and university campuses each year. … The majority of these are never reported in the published literature; consequently their strengths, limitations, and impact, along with any lessons learned by workshop facilitators, regrettably have not contributed to current knowledge about workshops. (p. 208)

> Most frequently, a participant questionnaire at the program's conclusion is the primary evidence collected to document program success. … Measures of participant satisfaction do not provide sufficient or direct evidence that a faculty development program has stimulated instructional improvement leading to enhanced student learning—an often stated goal of many programs. … One finds little evidence reported in the published literature that this type of data is collected to assess program impact. (pp. 216, 217)

As part of the current evaluation, SRI commissioned Dr. G. Roger Sell of the University of Northern Iowa to conduct a review of published literature regarding the impact of faculty enhancement (Sell, 1998). Only 12 studies regarding faculty met Sell's criteria for the review.[4] Although there were methodological differences across the studies and serious design flaws in some, the studies generally support the conventional wisdom that faculty professional development activities have positive impacts on teaching and learning, as well as on other behaviors of participants.

Teaching and instructional improvement. Participants themselves have reported improvements in their teaching because of workshops or related activities (Austin, 1992; Eble and McKeachie, 1985; Menges et al., 1988; Eison and Stevens, 1995). Students of participants also have rated faculty who participated in workshops as more effective teachers than comparable nonparticipant faculty or have indicated that participants' teaching improved after attending workshops (Annis, 1989; Boice, 1995; Hativa and Miron, 1990; Hoyt and Howard, 1978; however, see also Gibbs, Brown, and Keeley,

[4] To limit and focus the search and review process, Sell included only "research-based literature," which he defined as published studies that collected and used empirical data to assess workshop effectiveness. Purely descriptive or prescriptive reports were not included; neither were dissertations, unpublished professional meeting papers, occasional papers, project and technical reports, or ERIC documents.

1989). No studies reviewed by Sell relied on third parties' reports of the impact of workshops on faculty's subsequent teaching.[5]

Impact on student performance. The single study reviewed by Sell that examined the performance of participants' students in an objective way found a positive impact of faculty development activities. After participants attended a workshop focused on the teaching of a particular mathematics course, their students had higher rates of passing the course and higher final exam scores than students of a comparison group of faculty (Friedman and Stomper, 1983). Rates of passing the course and final exam scores also were higher than those of participants' own students before the training.

Nonclassroom activities. Several studies supported the notion that faculty development activities lead to increased productivity; Boice (1995), Eison and Stevens (1995), and Menges et al. (1988) reported positive effects of workshops on faculty research and scholarship. Collegial relationships and the involvement of faculty in departmental and institutional commitments also were found to be positively affected by workshops and related activities (Austin, 1992; Eble and McKeachie, 1985).

Findings from Westat's formative evaluation of the UFE program. A survey of UFE workshop participants conducted by Westat, Inc., in 1991 as part of a formative evaluation of the UFE program lent further credence to the positive impacts of workshops on participants' subsequent behaviors. Substantial majorities of respondents indicated that they had introduced new content into an existing undergraduate course or laboratory; acquired new equipment, materials, or computer software for undergraduate courses or laboratories; incorporated equipment, materials, or computer software into undergraduate courses or laboratories in a way they had not been used previously; modified teaching methods; and developed new undergraduate courses or laboratories (Westat, 1992).

In addition, majorities of participants indicated that, following the UFE workshop, they had participated in formal programs designed to develop curriculum or improve instruction, or had delivered a paper at a professional meeting or submitted an article to a

[5] Sell did review one study in which trained third-party observers judged the impact of an *entire semester* of training sessions on five faculty's subsequent teaching. The observers found that all five faculty implemented what they had learned during the training sessions.

professional journal; however, participants did not tend to attribute these behaviors strongly to the UFE program.

The UFE Evaluation Design and Methods

The Evaluation Design

This summative evaluation was designed to collect and interpret outcome and impact data on a programwide level to better understand the extent to which UFE as a whole met the objectives that NSF set out for the program. These objectives included the goals listed in the program announcement and in yearly directories of workshops. In addition, as the phasing in of the Government Performance and Results Act of 1993 (GPRA) became more influential in the Executive and Congressional examination of agency programs, NSF became more interested in knowing the program's broader impacts on undergraduate education.

To ascertain such impacts, SRI worked with NSF program and evaluation staff to develop a set of desired outcomes and indicators for the outcomes. Exhibit I-1 lists the outcomes (the complete set of outcomes and associated indicators is shown in Appendix A). Note that the outcomes focus on faculty behaviors, student performance, institutionalization of improved courses, etc., and are phrased in terms of desired states. Our evaluation was designed to measure movement toward those states by measuring *changes* in faculty behavior (i.e., development of new courses and/or revision of existing courses) that could be attributed to UFE.

Such changes are often measured by using a "pre-post" design or inferred from "comparison group" designs; however, neither of these was feasible for the present evaluation (see discussion below in "Limitations of the Evaluation Design"). Therefore, we observed workshops and asked participants directly in interviews and a telephone survey about changes they had made as a result of the workshops and about the consequences of these changes.

Exhibit I-1. Outcomes for the UFE Program

1. Faculty incorporate current and relevant content into their teaching, use state-of-the-art experimental techniques and technology, and apply best practices in instruction.
2. Undergraduate students, including those from underrepresented groups, gain proficiency in SMET, improve their attitudes toward SMET, and are prepared to apply SMET concepts to their lives.
3. Institutions offer SMET courses/labs for undergraduates that are state-of-the-art in their content and technology, incorporate best practices in their pedagogy, are accessible to all students, and are relevant to the real world.
4. SMET Faculty collaborate with one another and with other experts in their fields.
5. Reforms in undergraduate SMET courses are sustained.
6. Knowledge and skills from UFE workshops are disseminated widely.

The evaluation also was designed to examine the extent to which various factors were associated with participants' subsequent behaviors. Data for such analyses came from our survey, as well as from DUE's database.

The time frame covered by the quantitative data is 1991 through 1997. We began with 1991 because Westat's formative evaluation had examined workshops held in 1988-1990. We ended our coverage with 1997 because data from DUE were not available for workshops after that year, and because we chose 1996 and 1997 participants for our own survey. Qualitative data were collected in 1997 and 1998.

Limitations of the Evaluation Design

Our evaluation design has two principal limitations. The first is that it relies on participants' reports of changes in their own behavior and of their students' performance. The extent to which participants might be overly positive in their survey responses is unknown. Using third-party reports, rather than those of participants, would have been theoretically preferable; however, such a design was not feasible because it would have required visits to participants' classes before and after participants attended the UFE workshops. Because lists of participants were not available until after the UFE workshops, preworkshop observations were not possible. Second, even if lists had been obtainable, observations of a meaningful sample of participants would have been prohibitively costly.

Fortunately, two factors make reporting bias unlikely. First, participants had no vested interest in inflating the positive impact of the UFE program. In addition to the fact that the survey was anonymous, the UFE program had already been terminated by the time the survey was administered, so participants had nothing to gain. Second, some survey respondents had attended workshops run by PIs whose follow-up activities we had also visited. When we compared survey responses with interview data from participants at the same PIs' follow-up activities, we found them to be very similar regarding workshop impact.

A second limitation for the evaluation is created by the absence of a comparison group. Because of this absence, the evaluation can not tell us the extent to which nonparticipant faculty attended other types of workshops or no workshops at all, or the extent to which such faculty made the same types of changes as faculty who attended UFE workshops. However, because of a host of unmeasured factors, such as departments' push for reformed courses in particular areas and participants' motivation to attend the UFE workshops, finding a valid comparison group would have been an intractable task, and a poorly matched "comparison group" could result in misleading findings. Thus, we chose the simpler design.

Evaluation Methods

Our primary data collection methods were:

- A telephone survey of 1,118 faculty, and
- Site visits to 12 workshops or related activities and follow-up contact with PIs and/or participants at those activities.

Quantitative data collection. To allow faculty to have had sufficient time to develop and teach courses or otherwise implement new curricula, the telephone survey was administered to faculty 2 to 3 years after they had attended a UFE workshop. The survey instrument was developed to cover as many of the indicators shown in Appendix A as possible. In addition, the instrument contained questions regarding workshop activities and other variables that our preliminary qualitative data collection indicated might be associated with the outcomes. (A full discussion of survey data collection is presented in Appendix B. The survey instrument, annotated to show the linkage of each

survey item to the outcome/indicator it was intended to measure, is presented in Appendix C.)

Additional quantitative data were drawn from DUE's databases of UFE awards and PIs' responses to annual surveys. These data included such items as award amounts, numbers of applicants to workshops, numbers of participants at workshops, durations of workshops, etc.

Qualitative data collection. In 1997, we conducted a preliminary round of site visits to UFE follow-up activities or workshops that were in their second or third summer to gain insight for the development of outcomes and indicators and, thus, for our survey questionnaire. In 1998, a second round of site visits to actual workshops enabled us to observe firsthand their balance of activities and their quality. In-depth interviews conducted during these visits also provided us with a richer understanding of the workshops' leadership and participants than could be obtained through a survey. In 2000, we contacted PIs and selected participants by telephone and/or e-mail to learn about participants' postworkshop experiences. (Qualitative data collection methods are described more fully in Chapter III.)

Advisors for the Evaluation Design and Methods

An advisory committee consisting of five academic specialists in science, mathematics, or engineering and two specialists in the evaluation of educational programs was appointed to provide advice regarding the evaluation's design and methods (names and affiliations are shown in Appendix D). A full-day meeting was held, during which a tentative design and methods were presented to the committee for their discussion. The final evaluation design and methods took into account their comments.

Organization of the Report

Chapter II, *Description of the Workshops*, presents an overall picture of the size, scope, and disciplinary and thematic coverage of the UFE program. This chapter includes some information from our survey of participants, as well as summary data from DUE's annual surveys of Principal Investigators. Brief descriptions of selected workshop features and quotes from survey respondents are included here and in Chapters IV, V, and VI.

Chapter III, *Qualitative Findings*, opens by discussing our qualitative data collection methods, including how various workshops were chosen for site visits. It then presents some general observations about the workshops visited.

Chapter IV, *Quantitative Findings for Faculty and Institutions* focuses on what participants learned, how they applied that learning to developing or revising courses and/or programs of study for majors, and the extent to which new and/or revised courses were institutionalized. The chapter also discusses the workshops' impact on participants' own professional activities and the "ripple" effects of their sharing workshop information with colleagues. This chapter, as well as the two subsequent chapters, uses data from SRI's survey of participants.

Chapter V, *Quantitative Findings for Students*, presents estimates of how many students completed participants' new or revised courses, as well as participants' judgment of such students' performance.

Chapter VI, *Factors Associated with Workshops' Success*, reports on barriers that participants perceived to revising and/or developing courses. It then takes a multivariate look at the associations between workshops' characteristics and participants' likelihood of revising and/or developing courses. Lastly, it discusses participants' views on the importance of various workshop characteristics.

Chapter VII, *Conclusion*, presents observations and conclusions by SRI staff, based on the earlier chapters. References and appendices follow.