SUMMER 2010 • Volume 30, Number 3

Connie Kubo Della-Piana,
National Science Foundation
Gabriel M. Della-Piana,
University of Utah

CUR Focus on the Web

Evaluation to Improve Programs and Demonstrate the Value of a Research Experience for Undergraduate Students

If your undergraduate research program has been in operation long enough to be working reasonably well, you have a sense of what it is accomplishing or what you hope it will accomplish. However, to improve the program and demonstrate its worth to others you will need, at a minimum, credible documentation of how the program has been implemented (the extent to which it is operating as intended); evidence of program outcomes (the extent to which it achieves its objectives); and evidence of program impact (what would have happened in the absence of the program) (GAO 2005).

Although some evaluators (Chelimsky and Shadish, 1997) argue for empowering staff to evaluate their own programs, we strongly argue for the inclusion of an independent evaluator at some point in program development and operation. Nonetheless, a program director can adapt and implement some key strategies and methods to describe the program, pinpoint areas for improvement, and demonstrate the worth of the program. Multiple methods and multiple perspectives have become the standard in evaluation (American Evaluation Association, 2004). We draw on our own experience and unpublished reports to outline a set of strategies for evaluation, several methods, some noteworthy references, and descriptions of how the strategies can be used to produce information that meets the criteria of educational importance, potential for transfer to other settings, and some reduction of uncertainty concerning what makes the program worthy of support.

Ethical concerns over the collection of information from participants require that you seek a ruling on the effort to evaluate your program from your institution's Institutional Review Board (IRB). The IRB ruling has legal status and protects both participants and the program, so it is necessary to review your institution's IRB process and apply for IRB approval before you begin data collection.

The methods illustrated in this article can be used to evaluate undergraduate research experiences across disciplinary fields. The evaluation should integrate your knowledge of the undergraduate research experience, your understanding of the research program, and key issues and concerns unique to your institution and discipline(s). We highlight the following evaluation methods: the survey; focus groups; uses of institutional data on students, archival records, and student-assessment information; and historical tracing of outstanding accomplishments. Peer review is not described here, although we note it is one of the most common methods for experts in a field to review accomplishments, whether in arts, humanities, science, sports, or technology. Expert review draws on knowledge of the discipline and disciplinary standards that define "good work."

As interdisciplinary collaborations and research become increasingly prevalent, the difficulties of reviewing interdisciplinary projects using expert review become more evident (Langfeldt 2006). In studying specific cases of lack of agreement among reviewers, it has been found that developing clear procedural guidelines provides a structure for conducting a peer review and for documenting decisions grounded in reasoned disagreement. We conclude by describing a practical approach to evaluation that builds on key issues and concerns of the program and/or the higher-education context in which the program exists. We agree with Schacter (2002) that there are no ready-made, pre-packaged "tools" that an expert can hand a program director to guide the monitoring or evaluation of a specific program. Whether you begin with an outside evaluator or work on your own, what is essential is that you first develop a clear understanding of your program by conducting an analysis in which you articulate and describe its essential components and how it fits into the institution.

Some readers will be partially or well acquainted with evaluation and the methods described here. However, studies have found that in engineering, medicine, and other high-risk, complex performance areas, a set of reminders about the essential elements of a complex performance can reduce errors and highlight the omission of critical elements (Scriven 2010; Weick and Sutcliffe 2001). This article can serve both as a starter for evaluation and a checklist of reminders about what you know about designing and conducting an evaluation.

Some useful, complementary Web-based references include the W. K. Kellogg Foundation's *Evaluation Handbook*, which focuses on capacity building for the projects the foundation funds in health and community development; *My Environmental Education Resource Assistant* (MEERA), which

includes sample instruments and evaluations; the *Evaluation Toolkit* of the National Alliance for Media Art + Culture (NAMAC), which focuses on evaluation of youth programs but includes case studies that make it more broadly useful; the *Online Evaluation Resource Library* (OERL), which focuses on science, math, and technology; and the *American Evaluation Association's* Web site for professional evaluators, program directors, and members of the public interested in evaluation.

Describing Your Program

Before you evaluate your program, describe it and state what you most want to know about it. Answering the following questions will help in developing a description:

(1) What is the problem your program is addressing?

(2) What are the needs of the target population(s) served by the program?

(3) What are the program's goals?

(4) What is the program expected to accomplish?

(5) What are the core elements, processes, and activities that are designed and established to accomplish program goals?

(6) How does the program fit into the institution and the institution's goals?

(7) How, as a result of the program, will the target population(s) be changed?

(8) How, as a result of the program or its evaluation, will the program or institution be changed?

Your initial evaluation questions will grow out of this description, whether it is in narrative form or a graphic logic model or flowchart. For guides to describing your program in narrative form and identifying initial evaluation questions, see Della-Piana and Della-Piana (2008), and for a variety of graphic forms, see Owen and Rogers (1999) and an online handbook by W. K. Kellogg Foundation (2004).

In the following sections we describe methods to collect information about your program, and we conclude with a description of a practical approach to evaluation and guidance on communicating and reporting on the evaluation. As noted above, the evaluation methods we'll describe are the survey, focus groups, historical tracing, institutional data and archival records, and information derived from the use of assessment tools.

Surveys

Surveys are systematic ways of asking questions (and getting answers) using questionnaires and interviews (through face-to-face administration, by mail, through on-line services, or by telephone). See Fink (1995) on surveys and the GAO (1991) on the structured individual interview; Bickman and Rog (1998, Part III) cover both. Surveys can provide a broad picture of a situation for immediate use or point to areas of interest or clarification for later follow-up through in-depth study.

The design process. Following is a simplified guide to the design of a descriptive survey. Surveys also can be designed for other purposes (e.g., explanatory studies).

1. Examine a description of the program to identify what information you need. Identify the program's strengths and weaknesses, paying attention to obstacles or factors that facilitate the accomplishment of program goals.

2. Decide on the questions and types of question formats to ensure that you will gather the information of interest. Sources on the construction of questionnaire items are Bourque and Fielder (1995) and, for interview procedures, Frey and Oishi (1995). Examination of the following examples reveals the importance of careful consideration of the possibilities for gathering information using different types of items:

Open-ended items. These types of items allow you to tap into the perspectives of the respondents about an area of importance to the program or when it is important to uncover descriptions of personal experience that are not or cannot be categorized at the present time. For example, the following questionnaire item examines the nature of access to a mentor from the perspective of the respondent: "Describe the accessibility of your mentor. Provide a description of the characteristics of your access to your mentor that led you to rate his/her accessibility as high or low."

Note that one challenge to analyzing responses to open-ended questions is the development of a coding scheme that highlights and is sensitive to the perspectives of the respondents. See Fink (1995, 61-66) for thematic analysis and categorization.

Forced-choice response items. These types of items allow you to stipulate the type of response that is needed and the form in which the information can be analyzed and presented in standard ways (e.g., descriptive statistics). Analyses could include a description of responses to items by program participants or testing for statistical significance (e.g., responses from one

group are different from responses obtained from another group). For example, the following item examines alternative reasons for describing why access to a mentor is low:

Mark all the reasons that explain why you rated mentor accessibility low.

----- Mentor was not in the office during posted hours.

----- I didn't try more than once.

----- Mentor is a busy person and I didn't want to bother him/her.

Note that a challenge to the development of alternative choices for this format is constructing responses that reasonably mirror the experience of respondents (e.g., program participants). To develop relevant alternative responses to a forced-choice item, it is best to examine the answers given by a large representative group of respondents to an open-ended question (e.g., list all the reasons that explain why you rated mentor accessibility as low). Those reasons are then used to construct alternatives in a forced-choice format.

Agree-disagree response items. These types of items ask respondents to report on the degree to which they agree or disagree with a statement. The statements reflect common experiences associated with the program. For example, the following item examines the level of agreement/disagreement about the quality of the interaction between the respondent and his/her mentor: "On a scale of 1 to 5, in which 1 represents almost always, 3 represents 50 percent of the time, and 5 represents almost never, how often do you feel that INTERACTION between you and the mentor worked well?"

Note that respondents indicate level of agreement by selecting some point along the scale. What is known as "response set" is one challenge that needs to be seriously considered with these types of items. Respondents can develop a habit of responding at the middle of the scale to avoid making a judgment. Using a label for the middle position on the scale or using a four-point scale helps to avoid a response that does not discriminate between agree or disagree or mitigates against a respondent's avoidance of making a judgment.

3. Conduct a try-out or pilot test of your items. This procedure is designed to increase the likelihood that items on the questionnaire are clear and that the thinking behind the responses is comparable across respondents. The procedure includes having a small sample of participants in the program complete the questionnaire and then following that up with one-on-one interviews that focus on uncovering respondents' understanding of the items, as a way to discover poorly written items. Typical errors in question-writing that require revision are double-barreled items (asking two different things in one question), double negatives included in the wording (e.g., asking if the respondent favors or opposes not doing something), and use of ambiguous words (like "training" or "drop out" or "attended," without specifying definitions).

4. Administer the survey to a sample or the total group. If using a sample of program participants, make sure that the sample is representative of the total group in terms of important demographic variables (e.g., age, gender, race/ethnicity, time in the program). If you are conducting a face-to-face administration of the survey or a telephone interview, check to see that all questions are answered. If you are collecting information by mail or online, monitor and follow up with the respondents to ensure an adequate response rate. When reporting results, state the number of people who responded to each of the items. For ways to handle missing or incomplete responses, consult the references at the end of this article.

5. Analyze the results. An analysis of survey data might include a description of the demographics of the respondents, frequencies of responses to each of the items, relationships between some items (e.g., number of hours of study/work and test or performance outcomes; motivation and self-regulation of work and outcomes), and comparisons within the group of respondents (e.g., comparison of the frequencies of responses showing that males were more or less motivated than females; complex parts of the experience were easier for highly motivated participants than for less highly motivated participants). Note that we discussed analysis of open-ended items above.

Example of analysis of a survey's findings. Analysis of the data from one questionnaire survey revealed that all participants in a research team reported that they were involved in discussion and called on by the faculty mentor to elaborate or justify their problem-solving process or conclusions. When asked if the discussions were valuable in moving the research forward, a majority of respondents strongly agreed with the statement. However, a small number of respondents reported that they strongly disagreed with the statement. This pattern of responses was consistent across items that asked students about the

nature of the interaction during group work. This finding was supported by observations of team meetings and interviews with team members. This kind of confirmation provided support for the consistency of the pattern of responses found in the survey data. Data from three sources (e.g., survey, observation, and interview data) increase the credibility of information about the fidelity of the implementation of program activities. Additional examination of the data revealed that students who had less experience than the other students were minimally engaged by faculty mentors. This information led to a recommendation to modify faculty mentors' professional development to include strategies for monitoring themselves when they interacted with students during team meetings, in order to ensure engagement of students who are new to the research experience and/or have limited knowledge of the discipline.

Strengths and limitations of surveys. Surveys used for descriptive studies are best suited for obtaining a broad picture of a program or service, for using the information directly to ameliorate problems, or for identifying program components to study in greater depth. Moreover, the questions and item formats must be designed to counter self-reporting that may lead to biased reporting. Surveys are validated in part by use of other measures, such as interviews and observations. For useful survey-based instruments that assess the undergraduate research experience, see Lopatto (2004, 2007) and Seymour (2004).

The Focus Group

Focus-group methodology originated in sociology as a way to conduct group interviews (Merton, Fiske, and Kendall 1956). This methodology is popular in marketing and has been adapted as a tool for a range of situations in which a group has had a common experience. An updated source of method is Krueger and Casey (2009).

The design process. A focus group is designed around a common experience. The common experience in the focus group described below is the improvement of the quality of reporting by a research group on its collaborative research project. The steps included the following:

1. The focus. The project director decided that conducting a focus-group study would assist in discovering why students lacked confidence in accomplishing the project's goal of "developing written and oral communication skills in technical reporting." The choice of this focus was based on examination of project goals, the program's logic-map description, and results of a survey showing that only nine out of 21 students in three research teams reported confidence in achieving this goal.

2. The sample. Although there were existing research groups, it was decided that the focus-group study would include two focus groups with a mix of students from across the research teams (with 10 or 11 students per focus group). This arrangement minimized the strength of established patterns of interaction among working groups, as it was hypothesized that these patterns could interfere with the discussion of issues. Note that the composition of a focus group is based on the situation, cultural sensitivity, nature of the common experience, and the issues being addressed.

3. The arrangements. To set the stage for the study, focus-group leaders were selected, trained, and supplied with recording devices, space, nametags for participants, honoraria for participants, water or other beverages, and a procedural interview protocol. The protocol included a set of questions that were to be covered by the focus-group leaders. A key part of the training and protocol was managing the discussion and ensuring that all participants were included in the discussion. Note that keys to successful management of focus groups include training and practice in prompting participants to elaborate, clarify, or explain reasons behind their comments. Effective probes are to:

- Repeat the question when it seems that participants did not understand it or misinterpreted it or when they strayed off into other topics.
- Pause and nod expectantly, giving the respondent time to gather his/her thoughts.
- Repeat or paraphrase the respondent's comments, while nodding and writing down what was said.
- Encourage people to respond to comments by other participants, while preventing situations in which respondents talk over one another.

See the references at the end of the article for additional strategies for managing group interaction during focus-group discussion.

4. The analysis. There are several strategies for analyzing focus-group data. In this case, the leader conducted a thematic analysis of the comments made by participants in her focus

group. The analysis included a procedure for clustering statements that fit together around a common theme or problem. A reliability check was conducted by another focus-group leader. The check involved using the same procedure for clustering the same sample of comments and comparing the results of the two analyses.

Example of findings from the focus-group study. Once the focus-group data were analyzed using a procedure for conducting a thematic analysis, it was found that the university writing course was seen by most participants as providing guidance for writing essays, but not as providing guidance for writing a technical report describing research across a range of disciplines and modes of discourse, such as description (e.g., procedures or specific cases) and argument (e.g., making the case for a claim or assertion based on empirical evidence or findings). In addition, the analysis revealed that in several research groups, members did not facilitate participation by all members of the group in writing reports on the research. A technical-writing workshop was designed, in collaboration with the writing program, for students engaged in research. In addition, a simple guide was developed describing strategies for getting all members of the research team involved in producing a report. Hints were developed that helped students give feedback to each other and to encourage involvement and volunteering to write a section of the report. A checklist for feedback included the following ideas for responding to team members:

--- People won't be interested in this part.

--- People may not understand what this part means.

--- This part is really good.

--- The presentation of data is clear.

--- A reader could get lost here and lose track of what we are trying to get across.

--- Something is left out here.

--- I need to reread the whole section/report in order to understand the main point.

An unexpected finding pointed to another cause for lack of participation in writing a report. Several students reported that they were not interested in improving their technical-writing skills or in collaborating with others to write a report, because they believed that communicating well in oral and written forms was not an essential element for success in their chosen fields of interest. For instruments relevant to motivation, see Amabile et al. (1994) and Vallerand et al. (2003).

Strengths and limitations of focus groups. Focus groups are helpful in identifying and defining problems and strengths in the implementation of program activities and the completion of high-quality deliverables. While allowing for the cross-checking of perceptions across participants, it is costly to validate results of focus groups through follow-up analysis of documents, observations, or individual interviews. In addition, domination of the discussion by one participant can skew the information in a particular direction, rather than providing an understanding of the complexity of the common experience.

Historical Tracing of Current Accomplishments

Historical tracing (Ruegg and Jordan 2007) is a process that uses documents, interviews, and focus groups to trace and document how an outstanding outcome in a program has evolved. The method can be used to provide evidence of how changes in an organization, activity, individual performance, or product achieved its current level of quality and scope. The method provides insights into the process of innovation in the arts, science and technology, and the humanities. This kind of information can lead to ideas on how to improve organizational structures and procedures to facilitate future development of high-quality programs or innovations.

The design process. As with much innovation in science and technology, the humanities, or the arts, there is no one formula or set of steps. What might historical tracing reveal about how the Boston Pops and Tanglewood Festival chorus brought to perfection the performance of David Chase's arrangement of the "Twelve Days of Christmas," under the direction of Keith Lockhart? Or, what are the specific data sources and information that can be examined to trace how the treatment of sickle cell anemia evolved (Williams 2004)? There are no algorithms. However, a useful design process for historical tracing is presented here, in simplified form, that draws on the work of Ruegg and Jordan (2007) and can be used to trace exemplary practices in undergraduate research programs. The process:

1. Form a technical-review panel of experts to select the product, performance, or process to be studied (referred to as the "evaluand"). The panel provides background on the evaluand at the beginning of the study and critically reviews the findings at the end of the study.

2. Identify people, documents, and relevant references that provide a starting point for interviews and document analysis. Pinpoint key people, mechanisms, institutions, activities, and processes that seemed to play a role in the development of the evaluand (performance, product, or process).

3. Conduct interviews and document analysis with a focus on identifying planning processes, initial indicators of successes and failures that reveal progress, relationships between people or organizations that seem to contribute to accomplishments, the spread of effects, and evidence of comparative value over similar evaluands produced by others. Compare tentative findings to idealized versions of the process of producing creative works, innovations, or exemplary practices.

Insights from arts and humanities. Examination of a sample of abstracts from the National Conference on Undergraduate Research, NCUR22, in 2008 at Salisbury University reveals a set of research accomplishments that suggests possible indicators of innovation worthy of study through historical tracing. The key element that runs through many of the abstracts is how the arts and humanities interact with science and mathematics in ways corresponding to collaborations of mature researchers, as represented in Leach (2006). For example, psychologists in a project on choreography and cognition developed an interest in the physical control of dancers. Based on this interest, the psychologists reformulated their ideas about how and what one might research as scientific psychology. Or, during research on new materials (liquid crystal elastomers), the collaboration between scientist and artist moved from the artist's investigation of new material to the scientist's developing new ideas about materials from the artist's work. Many of the NCUR projects appear to have similar potential. Through examining these interdisciplinary activities, one might find something worth tracing to determine what brings the disciplines together.

Historical tracing of a biotechnology project. The following example is from a preliminary study that lays the groundwork for a more thorough and formal historical-tracing study. A biotechnology project that incorporated research experiences in a series of courses was identified by a funding agency as having outstanding results. A preliminary interview of the project director by a funding agency's program officer revealed convincing evidence of the project's effects on institutional infrastructure (obtaining funding for lab construction and a science building), high demand in both industry and university research labs for the program's graduates, and a publisher's interest in the program's materials (developed through collaboration between biotech industries and the community college-based program). Reasons given for recognition of the program highlighted the unique combination of the theory, research, and technical skills of the program's graduates, compared with four-year university counterparts who were strong on theory and research but who lacked technical skills and experience using biotech instrumentation. A formal, systematic study using the historical-tracing methodology could reveal the essential components that make course-based research experiences successful and highly regarded by industry and university-based research programs.

Strengths and limitations of historical tracing. To establish strong credibility for the findings, the method is time consuming, requires detailed documentation, and systematic and rigorous testing of alternative interpretations of the history of the case. However, with modest resources a convincing story can be told of the forces that helped influence program outcomes. If necessary, a more in-depth, follow-up case study along the lines proposed by Yin (2003) and Stake (1995) could validate findings from initial interviews and document analysis.

Institutional Data, Archival Records, and Data from Assessment Tools

Three strategies for documenting student outcomes are information collected by the institution for institutional compliance reporting or for institutional self-review; information collected by the program and stored in program records or a program database; and information derived from the use of assessment instruments that measure changes in knowledge or performance or that guide a critical review of products produced by students.

Institutional data. Information on student progress through degree programs forms the basis for supporting claims of an effective program, as indicated by increased retention in college and in the discipline, increased grade-point averages, increased graduation rates, and increased numbers of students involved in research accepted into graduate programs. Working with your college's institutional-research office to gather and analyze

student data requires resources, time, and compliance with privacy regulations. However, institutional data are certified, verifiable, linked to unique individuals, and can be compared to the general student population or a selected subset of the population. Be aware that there are variations among institutions and types of institutions in the resources allocated to collecting and making data on students accessible. Check with your institutional-research office to see if there are services to assist in the use of institutional data for purposes of program evaluation.

Archival program records. Standard information gathered about student participants should be archived in a program database. Knowledge about the program and the expected outcomes provide a framework for collecting essential information concerning student participants. Such information could include contact information, gender, race/ethnicity, discipline and/or major, class standing, grade-point average, future plans, and reasons for participating in the research experience or course. Based on program expectations, pre-research information can be collected, as well as post-experience information. For example, information could include completion rate, retention in the discipline, application to graduate school, transfer to a four-year program, or participation in another research experience in a department, college, university, or foundation-funded program. As a component of the research experience, students must provide specific information to the program. If a program database is to be maintained, procedures need to be in place for complying with privacy laws, securing the data, and limiting access to student information. Plans for collecting student data when the student completes or leaves the program must be developed and implemented within program resources.

Assessment data. Student-assessment data beyond the institutional database includes assessment of student outcomes using instruments such as departmental examinations or capstone projects, attitude surveys, inventories of specific concepts, tests assessing problem-solving or critical-thinking skills, performances (music, stage, presentations to professional associations), or portfolios of work in the arts or other fields. Some types of instruments are designed to measure well-defined bodies of knowledge or orientations. See Carlson (1985) for designs for assessment of instruction in English, social studies, and science. For the science and design of educational assessment, see Pellegrino, Chudowski, and Glaser (2001). For assessment of future plans, experiences with mentoring, and assessment of perceptions of summer research experiences and classroom research experiences, see Lopatto (2004, 2007).

Strengths and limitations. Access to institutional data requires a request to your institution's institutional-research office. Although certified, verified, and linked to individuals, special conditions may control access to and use of the information. In addition, institutional data is refreshed on a schedule, making replication of analyses challenging. Updating of a program database is time-consuming, and the database itself needs to be continually monitored to avoid degradation of the information, so check with your institution's IT group for advice. A thorough examination of an assessment tool's characteristics, current use, and interpretation of results is necessary to make an informed decision about its suitability for your program. For a guide to instrument review, see Nitko (undated).

An example of a longitudinal study. Research experiences across disciplines in arts and humanities, social science, physical and natural science, engineering, business, and education were included in a study of an 11-year-old collaboration (Foertsch, Alexander, and Penberthy 2000) involving summer research programs for minority-group students at a consortium of 15 Midwestern research universities (The Committee on Institution Cooperation). The study illustrates strategies for generating and using "institutional" data and records. The program provided guidance for conducting original research projects, weekly meetings with peers in workshops, networking opportunities and presentation of project reports in symposia, and strategies for letting undergraduates see what graduate students' life is like.

Major challenges to obtaining data for such studies were the logistics of tracking student progress, analyzing the Committee on Institutional Cooperation's exit survey for the program, and obtaining contact information to annually track student participants' progress toward a baccalaureate degree. Over 5,400 students participated in the program from 1986 to 1996. The study was conducted in 1996-1997, and successfully tracked 90 percent of the student participants. Findings revealed positive student outcomes associated with students' participation in an eight-week intensive program that was guided by a powerful conceptualization of the research experience and high levels of fidelity in the implementation of the program during the entire collaboration. Three quarters of the participants received their baccalaureate degrees and went on to graduate school or professional schools. It was found that student outcomes were

related to the effective implementation of mentoring roles, funding, recruiting, and follow-up efforts by departments.

A Practical Evaluation Approach Built Around a Set of Key Issues

A practical approach to evaluation draws on a diverse set of information that provides evidence to support assertions about the success, merit, and worth of an undergraduate research program and/or pinpoints areas for program improvement (Della-Piana and Della-Piana, 2008). The approach is designed to address the needs of a range of programs, from course-based to independent efforts at diverse institutions, and incorporates a variety of data-collection methods: surveys, interviews, focus groups, expert review, direct observation, participant observation, document analysis, and mining institutional and other types of archival records.

The keys to an effective evaluation are:

- A clear description of the undergraduate research program and its expected outcomes;
- Articulation of a set of issues/concerns and a set of evaluation questions directing the evaluation;
- Data collection and interpretation aligned with the goals of the program and the set of questions;
- Converging lines of evidence provided by a variety of data-collection methods; and
- Methods for addressing alternative explanations for the findings.

Evaluation should be conducted not only to demonstrate that a program is working, but also to improve the way it is working, and to inform decisions about future programs. A note of caution: There is no evaluation that answers every question with evidence that is credible for all stakeholders.

The design process. Steps in developing a practical evaluation include:

1. State the problem your program is addressing.

2. Describe your program or the logic of your program (See earlier information on describing your program) and state the purpose of the evaluation.

3. Decide what issues or concerns need to be addressed based on your knowledge of the program and the literature, what your program is designed to accomplish, your knowledge and experience with undergraduate research, and your understanding of your institution and your program's key internal and external stakeholders.

4. Generate evaluation questions based on the description of your program, and the purpose(s) of the evaluation (what you want to know and what you will do with the findings), carefully connecting them to the issues and/or concerns.

5. Prioritize your information needs by taking into account limited resources, including time, money, personnel, access to information, and expertise.

6. Specify expected outputs and outcomes (quantitative and/or qualitative indicators of success) that represent your program's accomplishments and could point to program limitations.

7. Select data-gathering tools that include multiple ways of collecting data and their sources that will provide evidence to support assertions of program success, merit, and areas for improvement.

8. Set out the evaluation framework and timeline. The framework links the evaluation questions, data, and data sources with a timeline and the responsible party or parties.

9. Develop a plan or approach for answering your questions. Select a standard research design that is described in a catalogue of quantitative and/or qualitative research methods. The three criteria for choice of design are credibility of information obtained (for a given audience and purpose), utility of information (for intended users of the information), and the economy, feasibility, and ethics of gathering information (given the available context and resources).

10. Ensure the collection of relevant and reliable data and analyze the data around the issues/concerns. Follow the standard "blueprint" of the research design, which includes steps for the collection, analysis, and interpretation of the data. Document all elements of the research protocol that describe the steps you will follow to collect, analyze, and interpret the data. Seek converging lines of evidence leading to the findings, while testing and countering alternative explanations for them.

11. Present findings in support of the assertions of program accomplishments and/or areas for program improvement. Discuss the limitations of the study.

12. Align the form and content of the reporting to the information needs of the targeted audience(s).

A humanities and social science study. A multiple-method study of a course-related research experience at a small liberal-arts college included a survey of 178 students across humanities and social-science disciplines. The findings focused on students' skills in using digital-information resources and how students conceptualize and operationalize research tasks, including writing four-to-six-page research reports (e.g., argument papers, literature reviews, or theory papers). Problems identified through survey data included: procrastination (73 percent of students); feeling overwhelmed by information (100 percent); difficulty in narrowing down topics (59 percent); and difficulty figuring out the professor's expectations for the work (48 percent). Confirmation of the survey's findings came from analysis of 30 assignment handouts by the professors (criteria for quality lacked clarity and detail). However, survey results indicated that 78 percent of the students found one-on-one sessions with the professor helpful and that 82 percent of the students found submitting drafts for comment was helpful. Analysis of two discussion groups revealed struggles with conceptualizing and narrowing down topics. This kind of survey study, used with other measures that allow deeper description and confirmation of findings, is useful for program improvement.

A computer-science team study. A program called the "cross-level research team experience in computer science" has been in operation for several years. Parts of the team experience were seen as not working well, and the program director wanted information to pinpoint areas that needed further development. Based on knowledge of the program and how it was working, there was some indication that several undergraduates felt a lack of knowledge concerning the discipline and the use of computers. While the program is built on a cooperative-learning model, several faculty mentors were more didactic (telling students what to do) than problem-oriented (having students and the faculty mentor take initiative as a team). A decision was made to address the following question in the evaluation: Why is the implementation of the undergraduate research-experience model based on a cooperative-learning paradigm only succeeding in highly monitored situations, as in faculty-development workshops, and not during the day-to-day research experience? The decision was to collect documents (reports of research progress from the previous semester), conduct focus groups with students (at three different stages of the semester), conduct individual interviews (with students participating in the research teams, faculty mentors for the teams, and other faculty mentors associated with the program), and assess students' perceptions of the research experience. A mixed-method approach was developed to examine the workload and responsibility of the faculty mentor, the individual student, and the team.

Findings revealed that the faculty mentors and students reverted back to an apprentice model for developing expertise, due in part to the institutional culture of the classroom that encourages and rewards such behavior in completing course-assigned projects. In addition, findings from the interviews with other faculty mentors suggested that overcoming this tendency was one of the hardest challenges in implementing a research program based on a cooperative-learning paradigm. These findings led to two recommendations for program improvement: pairing an experienced faculty mentor with a novice faculty mentor and developing a session during the faculty professional-development workshop dealing with implementation challenges that would discuss this specific issue at the time pairings of new faculty mentors occur.

Communicating and Reporting on Evaluation

Evaluations focused on the utilization of the evaluation process (Patton 2008) and findings argue that plans for ongoing use of the process, use of evaluation findings, and dissemination demand serious consideration. Rather than assuming that the evaluation process, findings, and recommendations will inform decision-making, utilization-focused evaluation makes the case that up-front planning for use of information throughout the evaluation facilitates drawing conclusions about a program. Planning for the use of information answers the questions "What happened?" "So what or why does it matter?" and "What should be done now?" The following questions will guide communication about and reporting on the evaluation:

- Who is the audience? Possible audiences are program staff, the students in the program, administrators who have responsibility for the program, funding agencies supporting the program, and colleagues in relevant professional associations. The form and content of reporting needs to be tailored to the information needs of your audience.
- What use of the information do you wish to facilitate? In general, the program staff's interest in the evaluation is directed at improving the program as it develops and

refines its activities and processes. At another level, staff members are interested in using information to demonstrate the program's worth to others, to justify continued or additional support, and to encourage others to implement it at other institutions or in other disciplines. Students may have an interest in how their comments were used and about the experience and perceptions of other students. Funding agencies will want to know that the outcomes are worth their investment or have the potential to be so.

- How can that use best be facilitated? Key components include articulating the purpose (e.g., program improvement, recruitment of students, accountability) and conceptualizing information "pathways." What is the purpose and how is the information expected to flow (e.g., through word of mouth, through formal institutional processes)? In what ways will the information be delivered (e.g., combination of graphic display and narrative, and combination of electronic, written, or face-to-face)? What venues are appropriate (e.g, professional meetings, public meetings, professional or popular publications)?

There are several strategies for reporting on an evaluation: the oral briefing, a written executive summary, a written technical report, and a public forum. The major take-away is that the kind of reporting is dependent on the purpose and the audience (Smith 1982; Torres, Preskill and Piontek 1996).

In conclusion, the purpose or set of purposes for conducting the evaluation guides and drives the evaluation. The strategies for the collection of information provide empirical evidence of the program's effectiveness/impact or lack of effectiveness/impact, and pinpoint areas for improvement through further study. Planning for the use of the evaluation process, findings, and recommendations assists in the identification and definition of information needs.

The ideas and opinions expressed in this article are those of the authors and do not represent the National Science Foundation or the University of Utah.

References

Amabile, Teresa M., Karl G. Hill, Beth A. Hennessey, and Elizabeth M. Tighe. 1994. The work preference inventory: Assessing intrinsic and extrinsic motivation orientation. *Journal of Personality and Social Psychology* 66 (5): 950-967.

American Evaluation Association. 2004. *American Evaluation Association Guiding Principles for Evaluators*. Fairhaven, Massachusetts: American Evaluation Association.

http://www.eval.org/Publications/GuidingPrinciples.asp. (accessed February 21, 2010).

Bourque, Linda B., and Eve P. Fielder. 1995. *How to conduct self-administered and mail surveys*. Thousand Oaks, CA: Sage.

Carlson, Sybil B. 1985. *Creative classroom testing: 10 designs for assessment and instruction*. Princeton, NJ: Educational Testing Service.

Della-Piana, Connie Kubo, and Gabriel Della-Piana. 2008. Evaluating the research experience. In *Affinity research group model: Creating and maintaining effective research teams*, ed. Ann Q. Gates, Steve Roach, Elsa Villa, Kerrie Kephart, Connie Della-Piana, and Gabriel Della-Piana, 47-52. Los Alamitos, CA: IEEE Computer Society.

Fink, Arlene. 1995. *The survey handbook*. Thousand Oaks, CA: Sage.

Foertsch, Julie, Blaine B. Alexander, and Deborah Penberthy. 2000. Summer research opportunity programs (SROPs) for minority undergraduates. *CUR Quarterly* 20 (3): 114-119.

Frey, James H., and Sabine Mertens Oishi. 1995. *How to conduct interviews by telephone and in person*. Thousand Oaks, CA: Sage.

GAO (Government Accounting Office). 2005. *Performance measurement and evaluation: Definitions and relationships.* GAO-05-739sp. Washington, DC: U.S. Government Accounting Office.

GAO (Government Accounting Office). 1991. *Using structured interviewing techniques*. PEMD 10.1.5. Washington, D.C.: U.S. Government Accounting Office.

GAO (Government Accounting Office). 1990. *Case study evaluations*. PEMD 10.1.9.Washington, DC: U.S. Government Accounting Office.

Krueger, Richard A., and Mary Anne Casey. 2009. *Focus groups: A practical guide for applied research*. Los Angeles, CA: Sage.

Langfeldt, Liv. 2006. The policy challenges of peer review: Managing bias, conflict of interests and interdisciplinary assessments. *Research Evaluation* 15: 31-41.

Leach, James. 2006. Extending Contexts, Making Possibilities: An Introduction to Evaluating the Projects. *LEONARDO* 39: 447-451.

Lopatto, David. 2007. Undergraduate research experiences support science career decisions and active learning. *CBE-Life Science Education* 6: 297-306. http://www.lifescied.org/cg/reprrint/6/4/297 (accessed February 19, 2010).

Lopatto, David. 2004. Survey of undergraduate research experiences (SURE): First findings". *Cell Biology Education* 3: 270-277. For instruments see: http://www.grinnell.edu/academic/psychology/faculty/dl/surecure (accessed January 12, 2010).

Merton, Robert K., Marjorie Fiske, and Patricia L. Kendall. 1956. *The focused interview*. Glencoe, Ill: Free Press.

Miles, Matthew B., and Miles Huberman. 1994. *Qualitative data analysis*. Thousand Oaks, CA: Sage.

My Environmental Education Evaluation Resource Assistant (MEERA). http://meera.snre.umich.edu

National Alliance for Media Art + Culture. http://www.namac.org/evalluation-toolkit.

Nitko, Anthony J. Using a mental measurements yearbooks review and other materials to evaluate a test. http://www.unl.edu/buros/bimm/html/lesson02.html (accessed December 14, 2009).

Online Evaluation Resource Library. http://oerl.sri.com

Owen, John M., and Patricia J. Rogers. 1999. *Program evaluation: Forms and approaches*. Thousand Oaks, CA: Sage.

Patton, Michael Quinn. 2008. *Utilization-focused evaluation,* 4th Edition. Los Angeles, CA: Sage.

Pellegrino, James W., Naomi Chudowsky, and Robert Glaser. 2001. *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.

Ruegg, Rosalie, and Gretchen Jordan. 2007. *Overview of evaluation methods for R & D programs.* Washington, DC: U.S. Department of Energy.

Schacter, Mark. 2002. *Practitioner's guide to measuring the performance of public programs*. Ottawa, Canada: Institute on Governance.

Scriven, Michael. 2010. Rethinking evaluation methodology. *Journal of MultiDisciplinary Evaluation* 6 (13): pp. i-ii. http://www.jmde.com/v6 (accessed February 9, 2010).

Seymour, Elaine, Anne-Barrie Hunter, Sandra L. Laursen, and Tracee Deantoni. 2004. Establishing the benefits of research experiences for undergraduates in the sciences: First findings from a three year study. *Science Education* 88: 493-534. http://www.wcer.wisc.eduj/salgains/instructor/SALGains.asp. (accessed January 19, 2010).

Smith, Nick L. (ed.). 1982. *Communication strategies in evaluation. New perspectives in evaluation.* Volume 3. Beverly Hills, CA: Sage.

Stake, Robert E. 1995. *The art of case study research.* Thousand Oaks, CA: Sage.

Torres, Rosalie T., Hallie S. Preskill, and Mary E. Piontek. 1996. *Evaluation strategies for communicating and reporting: Enhancing learning in organizations*. Thousand Oaks, CA: Sage.

Vallerand, Robert J., Sarah-Jeanne Salvy, Genevieve Mageau, Andrew Elliot, Pascale Denis, Frederic Grouzet, and Celine Blanchard. 2007. On the role of passion in performance. *Journal of Personality* 75: 505-533.

W. K. Kellogg Foundation Handbook. www.wkkf.org.

Weick, Karl. E., and Kathleen M. Sutcliffe. 2001. *Managing the unexpected: Assuring high performance in an age of complexity.* San Francisco, CA: Jossey-Bass.

Yin, Robert. 2003. *Case study research: Design and methods.* Thousand Oaks, CA: Sage.

Connie Kubo Della-Piana
Program Director
Division of Undergraduate Education
National Science Foundation
cdellapi@nsf.gov

Connie Kubo Della-Piana is a program director in the Division of Undergraduate Education at the National Science Foundation. Her current interests are in undergraduate education in science, technology, engineering, and mathematics; program evaluation; science and technology policy; and organizational studies. She has taught courses in program evaluation and organizational/interpersonal/group communication, as well as conducting short courses in evaluation at conferences. She has conducted evaluations or been

a consultant on evaluation for several organizations, including the BBC World Service Trust's Research and Learning Group, projects supported by the National Science Foundation, the National Institutes of Health's Minority Access to Research Careers program, the American Red Cross, and the Smithsonian Institution's Folk Art Festival.

Gabriel M. Della-Piana
Consultant on Program Design,
Development and Evaluation
Arlington, VA
dellapiana@aol.com

is professor emeritus in educational psychology at the University of Utah. He spent four years at the National Science Foundation as a program director in evaluation before becoming a private consultant in program design, development, and evaluation in Arlington, Virginia. He has taught program evaluation, instructional design, and literacy across disciplines. On leave from Utah, directed the Intercultural Reading and Language Program at the Northwest Regional Educational Laboratory in Portland, Oregon. He also has lectured on behavior theory for USAID in India; been director of research, curriculum, and evaluation at the El Paso Collaborative for Academic Excellence; and consulted on evaluation for the BBC World Service Trust's Research and Learning Group.