A LIVELY ELECTRONIC COMPENDIUM OF RESEARCH, NEWS, RESOURCES, AND OPINION

Astronomy Education Review

Volume 3, Oct 2004 - Apr 2005
Issue 2

# Finding the Forest Amid the Trees: Tools for Evaluating Astronomy Education and Public Outreach Projects

by **Janelle M. Bailey**
University of Arizona
**Timothy F. Slater**
University of Arizona
Received: 10/26/04, Revised: 02/16/05, Posted: 03/03/05

The Astronomy Education Review, Issue 2, Volume 3:47-60, 2005



### Abstract

The effective evaluation of educational projects is becoming increasingly important to funding agencies and to the individuals and organizations involved in the projects. This brief "how-to" guide provides an introductory description of the purpose and basic ideas of project evaluation, and uses authentic examples from four different astronomy and earth/space science projects to illustrate important ideas. Topics depicted include types of evaluation, selection of an evaluator, alignment of evaluation activities with project goals, and different methods of data collection in evaluation. A list of additional resources, including professional organizations and selected journals, is provided.

## 1. INTRODUCTION

## 1.1 What Is Project Evaluation?

Evaluation generally means to appraise or ascertain the value of something. Thus, *project* evaluation is a measure of the worth of a project, often for reasons of accountability. It is required of most externally funded service-oriented projects, regardless of discipline, and has similarities in purpose and scope to marketing research. Ultimately, we want to know if the project successfully reached the goals it set out to achieve. If not, what went wrong? How should the project be altered in the future to be more effective or successful?

Evaluation research as a widely recognized field of social science has grown dramatically since World War II, as major federal and private grants have required evidence of effectiveness for a variety of programs (Rossi, Lipsey, & Freeman 2004). Growth continued over the next several decades to the point that evaluation has become its own specialty field within the social sciences; the earliest journal, *Evaluation Review*, was first published in 1976. Greater and greater influence has been exerted by stakeholders and consumers--including public policymakers and administrators--over more recent years, and the field is now rapidly expanding beyond the social sciences (Rossi, Lipsey, & Freeman). Currently, the quality of the evaluation plan may be a deciding factor in awarding many education-related NASA and NSF grants.

Often the terms "assessment" and "evaluation" are used interchangeably. In some cases, this seems perfectly appropriate. In others, the distinction between the two is paramount. For the purposes of this article, evaluation is the process by which data are collected and a judgment is made based upon these data regarding the effectiveness or worth of an educational project. Assessment can then describe the collection of data (sometimes for the purpose of judgment or accountability, but not exclusively) from an individual student, for example.

## 1.2 Why Do It?

As educators, we are all familiar with the evaluation of our students (i.e., assigning their grades) and their evaluation of us (e.g., end-of-course evaluations). However, many educational projects, curricular materials, and workshops are evaluated much less systematically than our students' grades, often relying on isolated anecdotal evidence or simply our gut feelings about how something worked in the classroom. By taking a formal, systematic approach to project evaluation, we can improve the effectiveness and credibility of our projects through the use of proven research-based methods in education. As we also know from our students, the things we do in our classrooms may result in understandings very different from what we anticipate; evaluation can serve to document such unexpected or unintended outcomes in a systematic manner. Moreover, an emphasis on high-quality project evaluation at the outset of project planning can dramatically improve the chances that a proposal will be funded in the first place.

As research scientists become increasingly involved in education and public outreach, it is important to try to learn about evaluation from those who have gone before. Now more than ever, principal investigators and program managers are being held accountable for the effectiveness of their projects and any resulting products. Some education projects have experienced great success. Others have very publicly and dramatically failed, while most simply fade into obscurity when the funding disappears. Nearly all funding agencies now require that an evaluation component be built into projects from the outset; in fact, most large NSF-funded education projects allocate 5%-10% of total direct costs for evaluation expenses (Frechtling 2002). So important is a comprehensive evaluation component that the NSF has published handbooks (Frechtling, 1995, 2002; Frechtling & Sharp, 1997; Frechtling et al. 1993; see Note 1) and conducted workshops to guide scientists in the evaluation of their education projects.

The handbooks produced by Frechtling and colleagues (Frechtling 1995, 2002; Frechtling & Sharp 1997; Frechtling et al. 1993) are designed to acquaint researchers with commonly practiced evaluation techniques. Although they draw from a wide body of evaluation literature and from experience, these handbooks emphasize evaluation needs and techniques specific to the NSF, and so may be especially useful for NSF-funded projects. In the case of a NASA IDEAS grant, the *Elements of a Program Evaluation* Web site may offer program-specific suggestions for evaluation (Space Telescope Science

Institute 2005).

In this article, we hope to help readers understand some of the fundamental components of project evaluation so that they can effectively incorporate it into their education projects. To do so, we will describe project evaluation in general, and use examples from our evaluation efforts for several different projects to illustrate specific aspects.

# 2. EVALUATION AS PART OF AN ASTRONOMY EDUCATION OR PUBLIC OUTREACH PROJECT: A "HOW-TO" GUIDE WITH EXAMPLES

## 2.1 Background of the Example Projects

Throughout this article, we illustrate major issues by infusing examples from real-life project evaluations into the text. It is not our intent to present comprehensive evaluation results from any of these projects or to critique them publicly; rather, we aim to provide examples from multiple sources that exemplify the points being made. One or both of the authors were involved with the evaluation of each of these projects. The Toward Other Planetary Systems (TOPS) and Authentic Research on Variable Stars (VS) projects are professional development workshops for teachers, the former a three-week residential workshop focusing on the integration of astronomy (including archaeoastronomy and astrobiology) into the classroom, and the latter a 10-week part-time training program in the use of a research protocol for investigating variable star data available on the Internet. The University of Hawai'i's Research Experiences for Undergradutes (UH REU) program is a 10-week summer research program in astronomy. Finally, the Princeton Earth Physics Project (PEPP) created a network of seismometers in middle and high school classrooms across the country. Additional information about each of these programs can be found in Notes 2-5.

## 2.2 Selecting an Evaluator

The selection of an evaluator or evaluation team is not a simple matter. In general, evaluators work for the project investigators or directors and not for the funding agency, and as such, it is the project directors who decide what information is released to the funding agency or into the public domain. As Frechtling (2002) described, it is desirable to enlist the services of an evaluator from the beginning of the project, although this may not always be possible. Although the evaluator can be internal to the project investigators' department, it is often preferable to enlist someone from outside the project. This could be a colleague from another department at the same institution (such as education or psychology departments), or someone from a professional evaluation company. The evaluator should have a good understanding of the project's goals and objectives and be accepting of them without introducing bias. It is also preferable for the evaluator to have at least a basic understanding of the scientific content underlying the project.

## 2.3 Types of Evaluation

The appropriate use of project evaluation is analogous to that of the selection of suitable assessment techniques in our classrooms (Brissenden et al. 2002). As described in greater detail by Hannah (1996), evaluators generally use three types of evaluation. Other evaluative studies can be used (see Rossi, Lipsey, & Freeman 2004, pages 62-65, for a definitional list), but those described below tend to be the most common. The choice and depth of evaluation type depends largely on the project under scrutiny and the goals of the project directors and evaluators. In the ideal case, multiple types of evaluation would be included over the lifetime of the project.

A *planning* evaluation serves to assess a project in its planning stages--in essence, to align a project's goals, objectives, processes, and timelines. This helps to focus project goals and, when used early enough in the proposal development process, can uncover weaknesses or inconsistencies in proposed programs before they are submitted for merit review. A *needs assessment*, a systematic study of the needs of the program and its potential consumers, may be included as part of the planning evaluation.

A *formative* evaluation assesses a project while in progress, making suggestions for change and further evaluating any midproject alterations that are implemented. It is ongoing throughout the project and may be iterative in nature. In contrast, a *summative* evaluation looks at a project only upon its completion. This is performed with the intent of making a final judgment about the level of success of the project--for example, by determining if its goals and objectives were met.

## 2.4 The Importance of Project Goals

Just as in a well-designed astronomy course, the most successful astronomy education and public outreach projects are those that have plainly articulated goals and a clear path to achieving them. Remember that the purpose of an evaluation, especially the summative, is to determine whether the project goals have been met. The presence of well-defined goals was probably part of what helped the project receive funding; don't lose sight of those goals once this happens.

It is also important to make explicit all of the goals that will be evaluated. Often, we have implicit goals--an example is to improve a participant's attitude toward astronomy--that go unstated but are assumed by the persons involved in the project. These implicit goals may then surface during a project evaluation without ever having been addressed by the project in a measurable way.

### *2.4.1 Types of Project Goals*

Science education projects are frequently built upon the following types of goals: (1) cognitive goals designed to increase participants' knowledge of science concepts or to improve their scientific inquiry skills; (2) affective goals aimed at enhancing participants' attitudes, values, and interests (defined further in section 2.6.3) in science; and (3) product creation and dissemination goals for classroom-tested instructional materials or techniques. Projects may focus on only one of these types of goals, or may have all three types incorporated. The VS project, for example, has all three types of goals at varying levels. The cognitive goal is that teachers (and eventually their students) who used the research protocol should demonstrate increases in knowledge of variable stars. Generating participants' interest in using the protocol is related to affective goals, and the widespread distribution of the student research protocol and accompanying guide book for teachers is a clear example of a product creation and dissemination goal.

### *2.4.2 Goals versus Objectives*

Instructors can generally attest that having overarching goals for their class is important, but also can recognize that broad goals are not easily measured. Instead, we often break down our instructional goals into smaller chunks--objectives or measurable outcomes--so that we might more easily and effectively establish the extent to which our goals have been met. For example, if a project (or classroom) goal is to significantly increase participant interest in cosmology, objectives might include increased scores on an interest survey when measured over the course through pretests and posttests, increased time spent with media related to cosmology (such as choosing to read magazine articles or watch science-based television programs), or the voluntary selection of cosmology as a topic for a paper or project. In much in the same way that scientists use different data sources to establish the validity of their conclusions, multiple objectives often are needed to determine if a broad goal has been met, providing a way to "triangulate" the evaluation data through a variety of sources.

### *2.4.3 An Example of Explicit and Implicit Project Goals*

In the PEPP evaluation, the original goal of the project (established by a different principal investigator from the one who contracted the evaluation efforts) was simply to create a network of working seismometers located within high schools across the country. This is actually quite simple to evaluate by counting the number of active seismometers within the network. However, implicit goals that were eventually made explicit for the purposes of evaluation included, for example, the use of the real science data in the classroom and affective gains toward learning about earthquakes. The summative evaluation for PEPP met with many challenges as a result of the low number of project goals explicit in the initial design of the project and the variety of implicit goals that became apparent during the one-year evaluation period.

## 2.5 The Evaluation Matrix: Aligning Goals with Measurable Outcomes

A common format to summarize evaluation plans is a matrix that relates the various goals of a project to the evaluation procedures and objectives to be met (the evaluation matrix for the VS project is provided in Table 1 as an example). Rows of the matrix describe the specific project goals and outcomes; columns indicate project activities, assessment data sources and analysis strategies, and performance indicators of success for each of the listed goals. As such, the evaluation matrix provides a structured approach for the external project evaluation, and clearly specifies how the evaluation is related to the project’s desired outcomes and definitions of success. The development of such an evaluation matrix during planning is essential to making project goals and activities more explicit for the evaluators, program directors, and proposal reviewers. When clearly laid out, the matrix directs the work of the evaluator and also provides merit reviewers a succinct and clear overview of a proposed project, its planned evaluation, and its indicators of success.

**Table 1.** Evaluation Matrix for the Variable Star (VS) Project

| Project Goals and Outcomes | Project Activities | Assessment Data Sources and Analysis Strategies | Indicators of Success |
|---|---|---|---|
| To increase participants' knowledge of how variable stars are studied. | Provide participants with 5 hours of lecture, 10 hours of hands-on activities, and 2 hours of video. | 1. Pre-post multiple-choice knowledge test.<br>2. Individual interviews with participants. | 1. Statistically significant test score increases.<br>2. 80% of participants can correctly describe how variables are studied. |
| To improve participants' attitudes about, value for, and interest in astronomy as a lifelong learning activity. | Provide and moderate online discussion groups about recent news releases. | 1. Pre-post 1-5 scale attitude survey.<br>2. Analysis of electronic discussion transcripts.<br>3. Open-ended survey about lifelong learning. | 1. Statistically significant increases on attitude survey.<br>2. 50% of participants are active discussants.<br>3. Survey demonstrates that 80% of participants have active lifelong learning interests in astronomy. |
| To increase the amount of class time devoted to contemporary astronomy concepts in teacher-participants' classrooms. | Conduct 15 hours of teacher workshops on how to use "Hands-On Astrophysics" (HOA) curriculum package. | 1. Pre-post survey asking participants how many classroom hours are devoted to astronomy.<br>2. Focus group interviews. | 1. Statistically significant increases in number of hours devoted to astronomy.<br>2. Interview transcripts show that 80% will use the HOA materials next year. |

## 2.6 Evaluation Techniques

### *2.6.1 Quantitative and Qualitative Methods*

Project evaluation, like educational research in general, can encompass a variety of quantitative and qualitative data collection strategies. Systematic qualitative methods might include, but are not limited to, repeated classroom observations, collections of participants' work, or clinical and group interviews of participants. While such techniques are ideally suited for gathering insightful anecdotal evidence or describing a single event or participant, they tend to work best for time-intensive case studies. However, this is usually insufficient for making wide generalizations that can be applied to other projects or participants. Alternatively, quantitative methods--surveys, for example--allow more room for statistical analysis and perhaps wider generalizations. Unfortunately, these methods are also limited in scope and might not thoroughly describe a given situation if the evaluator is not aware of the nuances of the project or specific scientific concepts.

The most useful evaluation studies use a combination of quantitative and qualitative data sources. In particular, results obtained from quantitative instruments need to be validated qualitatively using individual or group interviews (or other qualitative techniques) with participants. One approach is to use qualitative methods to inform the design of quantitative instruments, later revisiting qualitative methods to support the analysis and interpretations of the data. Some evaluators accomplish this in part by presenting the qualitative data to project participants and asking them to help interpret the results. The evaluation results that are the most convincing employ a triangulated multiple data source approach to assess stated project goals with both quantitative and qualitative collection strategies. Such triangulation of data also helps reduce potential weaknesses in a single data source.

### *2.6.2 Measuring Knowledge and Skills*

Changes in knowledge (the cognitive domain) among participants are commonly measured using pretests and posttests that cover either a narrow range of content knowledge specific to the project, or broader science content. Only a few national instruments exist with large comparison databases to measure broad understanding of astronomy and earth/space science concepts, and none is yet in widespread use. These instruments include the Astronomy Diagnostic Test (Hufnagel et al. 2000), the AGI/NSTA Earth Science Examination (Callister et al. 1988), the Earth Science "Literacy Test" (DeLaughter et al. 1998), and the Nature of Science Survey (Libarkin 2001). The Astronomy Diagnostic Test was used in pre/posttest administrations for the TOPS workshop nearly every summer in order to assess what astronomy content knowledge was gained by the participants over the three-week period. However, it is much more common for evaluators, or even the project investigators, to develop their own pretest/posttest knowledge surveys that are responsive to the emphases of individual projects. Usually multiple-choice in format, these surveys are time consuming to develop and suffer from many of the weaknesses of such instruments outlined elsewhere (Astwood & Slater 1997).

Experience suggests that participant skills (unlike content knowledge) that have improved as a result of an activity or project can be measured reasonably reliably by self-report if participants do not have a reason to inflate the results. Alternatively, participants can be observed and the important components of a performed skill can be rated by the evaluator using an observation checklist. In the UH REU evaluation, interviews of participants at both the beginning and end of the summer term often included questions about their observational or data reduction skills.

### *2.6.3 Measuring Attitudes, Values, and Interest*

Whereas the cognitive domain focuses on what participants understand and can apply, the affective domain embraces attitudes, values, and interests. In the language of education, *attitudes* are the extent to which participants like or enjoy something, *values* are the degrees to which participants think that something is important to engage in, and participants’ *interests* are things worthy of allocating time to

(Anderson 1981). These facets of the affective domain may be related to, but quite different from, participants' satisfaction with the program itself. The affective domain is most commonly evaluated by asking participants to self-report their feelings using pretest and posttest Likert scale surveys. Participants are presented with a statement and asked to respond on a scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). As an example, Table 2 shows the survey that was administered after the 2003 TOPS workshop. Typically, mean scores and standard deviations are calculated for each item, and a *t* test of significance is calculated for a given response to judge the meaningfulness of any observed change between preproject and postproject administrations.

**Table 2.** TOPS 2003 Postworkshop Questionnaire

| Name (optional) ______________________________ | Strongly Disagree | Disagree | Neutral or Don't Know | Agree | Strongly Agree |
|---|---|---|---|---|---|
| 1. I have the knowledge to incorporate astronomy activities and/or content into my class. | 1 | 2 | 3 | 4 | 5 |
| 2. I have adequate resources to incorporate astronomy activities and/or content into my class. | 1 | 2 | 3 | 4 | 5 |
| 3. I have the skills to incorporate astronomy activities and/or content into my class. | 1 | 2 | 3 | 4 | 5 |
| 4. I feel confident in my abilities to incorporate astronomy activities and/or content into my class. | 1 | 2 | 3 | 4 | 5 |
| 5. I am interested in incorporating astronomy activities and/or content into my class. | 1 | 2 | 3 | 4 | 5 |
| 6. I have the knowledge to guide students in a remote observing project. | 1 | 2 | 3 | 4 | 5 |
| 7. I have adequate resources to guide students in a remote observing project. | 1 | 2 | 3 | 4 | 5 |
| 8. I have the skills to guide students in a remote observing project. | 1 | 2 | 3 | 4 | 5 |
| 9. I feel confident in my abilities to guide students in a remote observing project. | 1 | 2 | 3 | 4 | 5 |
| 10. I am interested in guiding students in a remote observing project. | 1 | 2 | 3 | 4 | 5 |
| 11. I have the knowledge to present a talk or lead an activity at a workshop (such as NSTA or HSTA).* | 1 | 2 | 3 | 4 | 5 |
| 12. I have adequate resources to present a talk or lead an activity at a workshop (such as NSTA or HSTA). | 1 | 2 | 3 | 4 | 5 |
| 13. I have the skills to present a talk or lead an activity at a workshop (such as NSTA or HSTA). | 1 | 2 | 3 | 4 | 5 |
| 14. I feel confident in my abilities to present a talk or lead an activity at a workshop (such as NSTA or HSTA). | 1 | 2 | 3 | 4 | 5 |
| 15. I am interested in presenting a talk or leading an activity at a workshop (such as NSTA or HSTA). | 1 | 2 | 3 | 4 | 5 |
| 16. I have the knowledge to organize a professional development workshop for other teachers. | 1 | 2 | 3 | 4 | 5 |
| 17. I have adequate resources to organize a professional development workshop for other teachers. | 1 | 2 | 3 | 4 | 5 |
| 18. I have the skills to organize a professional development workshop for other teachers. | 1 | 2 | 3 | 4 | 5 |
| 19. I feel confident in my abilities to organize a professional development workshop for other teachers. | 1 | 2 | 3 | 4 | 5 |
| 20. I am interested in organizing a professional development workshop for other teachers. | 1 | 2 | 3 | 4 | 5 |

* NSTA = National Science Teachers Association; HSTA = Hawaii Science Teachers Association.

One concern about surveys that use such a scale is the variability across participants; what one person agrees with, another might strongly agree with. Likewise, an individual might feel different from one administration to another. Often results are binned into *positive* (e.g., responses 5 and 4), *neutral* (3), and *negative* (2 and 1) categories to help alleviate the potential differences between participants or over time. These are questions of validity and reliability, respectively. More information on how to design validated, reliable surveys can be found in any number of references on testing and measurement, such as Campbell & Stanley (1963). Other data collection methods, especially qualitative methods such as interviews, might more accurately represent participants' true feelings in the affective domain.

### *2.6.4 Product Creation and Dissemination*

Science education projects often include the development of curriculum products. It is important that these products be thoroughly evaluated (including classroom testing) to judge their ease of use, appropriateness, value, and effectiveness before they are widely disseminated. The measurement of these attributes is often tied to the previously described measures of the cognitive and affective domains. Similarly, a careful plan at the beginning of the project for how and how widely the materials will be disseminated and adopted may point to fixable weaknesses and thus increase the later success of the dissemination plan. Dissemination of the VS research protocol is one of that project's three primary goals; as such, the tracking of its distribution and use will be an important part of the evaluation efforts. Continued contact with current and future teachers will be paramount to the success of this part of the project evaluation.

## 2.7 Using Evaluation Findings to Improve Education and Outreach Efforts

The uses of planning and formative project evaluations are fairly straightforward. Because they occur before or during a project, changes can be made based upon evaluation results in order to improve the project if the project leadership values and actively participates in the evaluation process. During the TOPS program, daily forms were collected with evaluation comments on each session offered. The project directors could use this information, compiled in a report submitted at the end of the summer, to determine which sessions to repeat (or not) during future summer workshops. Also, if immediate needs or concerns were provided by participants on a daily evaluation form, project staff could address them within one to two days.

Summative evaluations have their own place in improving education and outreach efforts, though their intrinsic value may not be as obvious as that of formative evaluation. Lessons can be learned both about the project itself (e.g., how to best achieve certain goals, logistics that need to be reconsidered) and the evaluation process. Sometimes we find that our data collection was not the best for the objective or goal described; this can be fixed in future projects and their evaluations.

## 2.8 Common Evaluation Pitfalls to Avoid

The most tenacious problem with evaluation efforts is the self-selection effect shown throughout the process. Especially in the case of teacher workshops, participants who are the most willing to participate in evaluation activities (such as interviews or the completion of surveys) are often those most motivated to succeed, either intrinsically or as a result of the project with which they are involved. This self-selection effect can skew results in a positive direction, making the project appear quite successful when in reality, a portion of the participants (and their data) are not represented. In the case of PEPP, those teachers who were active and excited about the program were happy to contribute their time to evaluation efforts. However, several of the original teachers had moved on to new jobs or simply had not continued their participation with the seismic network. Not surprisingly, these teachers often could not be contacted to determine why their participation had lapsed and what could be done to alter the situation.

An additional challenge faced by the PEPP evaluation team was its late involvement in the project. PEPP had been funded at some level since 1993, but the formal evaluation efforts did not begin until 2000, largely as a result of changes in funding requirements as they related to evaluation. In this case, a summative evaluation suffers from the inability to investigate or demonstrate the potential changes in the cognitive and affective domains as a result of the project; it is impossible to collect pretest data after the fact.

As mentioned earlier, sometimes we don't realize until after the analysis is complete that a data collection strategy is less effective than we would like. Early during the TOPS evaluation, we used the Astronomy Diagnostic Test (Hufnagel et al. 2000; Zeilik 2003) to ascertain cognitive growth as a result of participation in the workshop. However, upon analysis, it was noted that many of the topics included on the ADT had not been addressed in the TOPS workshop; likewise, many astrobiology and archaeoastronomy topics were covered in the workshop but not assessed in the ADT. The instrument was subsequently changed to more accurately reflect the topics emphasized in the workshop, incorporating questions from the ADT and other sources. Any data-driven claims about how the workshop impacted repeat participants were, of course, compromised by this change.

The collection of longitudinal data (tracking participants months or even years after their initial involvement in a project) is one of the most valuable and most difficult of all evaluation tasks. Funding for such evaluation efforts is rarely available, and often deadlines for reports to funding agencies preclude the collection of long-term data. However, knowing the influence of a project "down the road" (what is known as an *impact* evaluation) can be beneficial for similar or extension projects of the future.

# 3. SUMMARY

As scientists become more involved in education and outreach efforts, it is vital that they also adopt the practice of actively and thoroughly evaluating the effectiveness of those efforts. This should be pursued with the same scholarly perspective that is fundamental to their scientific endeavors. This article provides an introduction to the common tools of project evaluation, describing several types of evaluation, methods of data collection, and "lessons learned" from a number of authentic project evaluation efforts. With some guidance and practice, project evaluation can improve the quality and effectiveness of education and outreach proposals and projects.

## Acknowledgments

The evaluations described in this article were supported in part by NSF REC-9903153 and NSF TE-9731083 (PEPP); NSF TE-9731083 (TOPS); NSF AST-9987896 (UH REU); and an Arizona Workforce Development grant (VS). The statements, conclusions, and recommendations represent the views of the authors and not necessarily the National Science Foundation, nor the projects' investigators. We extend our appreciation to Michelle Hall-Wallace, Kirstin Bittel, Rachel Hughes, and Marti Lindsey for their helpful comments on earlier drafts.


## Resources

**Professional Organizations**
American Evaluation Association: http://www.eval.org
Association for Public Policy Analysis and Management: http://www.appam.org/
American Educational Research Association (Evaluation Division): http://www.aera.net
Canadian Evaluation Association: http://www.evaluationcanada.ca
Australasian Evaluation Society: http://www.parklane.com.au/aes
European Evaluation Society: http://www.europeanevaluation.org

**Sample of Evaluation Journals**
*Evaluation Review: A Journal of Applied Social Research* (Sage Publications)
*American Journal of Evaluation* (formerly *Evaluation Practice*; JAI Press)
*New Directions for Evaluation* (Jossey-Bass)
*Evaluation: The International Journal of Theory, Research, and Practice* (Sage Publications)
*Assessment and Evaluation in Higher Education* (Carfax Publishing Ltd.)

## Notes

**1.The National Science Foundation evaluation handbooks** by Frechtling and colleagues are freely available through the National Science Foundation's Web site: http://www.nsf.gov/publications/. The full reference citations are given below.

**2. Toward Other Planetary Systems (TOPS) Teacher Enhancement Workshop**
Focused on high school science teachers from the Hawai'ian and Pacific Islands, the Toward Other Planetary Systems (TOPS) Teacher Enhancement workshop used the theme of cutting-edge astronomical research to motivate and train high school teachers in order to enhance the quantity and quality of astronomy instruction in their classrooms. The first four years of the program were funded by NASA, with an additional five years sponsored by the NSF. Each year's workshop lasted for three weeks and included 20-30 teacher participants and approximately 10 staff members. Participants implemented astronomy lessons in their classrooms during the following year and shared their experiences with colleagues through professional development venues. They were also required to submit to the principal investigator a portfolio that showcased these activities and provided an avenue for reflection on their experiences.

**3. University of Hawai'i's Research Experiences for Undergraduates in Astronomy (UH REU)**
The University of Hawai'i's Research Experiences for Undergraduates (UH REU) program in astronomy allows six to eight undergraduate students to engage in research over 10 weeks every summer. Students have the opportunity to experience a variety of research techniques, including observing, reducing data,

analyzing data, building instruments, and programming code. Often, such experiences help focus a student's interests regarding his or her future educational and career path as it relates to astronomy. A student is partnered with a professional astronomer on a short-term research project. At the end of the program, students create poster and oral presentations about their research to put on for their peers; many also have the opportunity to publish or present their results at national conferences. Sponsored by the NSF, the UH REU program is one of more than 20 in astronomy nationwide, and REU programs exist in several other disciplines as well.

**4. Princeton Earth Physics Project (PEPP)**

The Princeton Earth Physics Project (PEPP) was an eight-year effort to develop a network of seismometers in high school classrooms across the country. Started in 1993, the original goal of PEPP was to develop the seismic network and use it to collect data from a variety of locations. However, late in the program, PEPP expanded its goals to include some professional development for teachers and a limited amount of curriculum development. More than 70 seismometers were installed in different high schools and middle schools across the country. They were operated by classroom teachers and their students under the guidance of university faculty who served as network hosts; physics and earth science teachers were the primary participants. Participating schools included public and private middle and high schools in both urban and rural areas, and were usually located within a few hours of their university network host.

**5. Authentic Research on Variable Stars (VS)**

Most professional development programs for teachers last only a few days, or at best two to three weeks. The intensity of such workshops is high, and little time is available for deep reflection or creation of implementation plans for the new ideas learned. Instead of this more typical format, the Authentic Research on Variable Stars (VS) project consists of an extended (10-week) part-time summer workshop for science teachers already in residence at the University of Arizona. Nine teacher participants worked in 2004 to develop and test a research protocol for secondary students to use to conduct research on variable star data available through the Internet; 10 students will work during 2005 to test the protocol. Such a protocol is suitable for student use in science fairs or independent study projects. The VS project, funded by an Arizona Workforce Development grant, has three major goals: to develop the aforementioned research protocol, to evaluate an extended professional development model, and to disseminate a guide to support teachers who supervise students using the research protocol.

Please address correspondence to Janelle Bailey, 520-626-9480, Fax 520-621-1532; jbailey@as.arizona.edu

## References

Anderson, L. W. 1981, *Assessing Affective Characteristics in the Schools*, Boston: Allyn & Bacon.

Astwood, P. M., & Slater, T. F. 1997, The Effectiveness and Management of Portfolio Assessment in Large-Enrollment Courses, *Journal of Geoscience Education*, 45, 238.

Brissenden, G., Slater, T. F., Mathieu, R. D., & the National Institute for Science Education (NISE) College Level-One Team. 2002, The Role of Assessment in the Development of the College Introductory Course: A 'How-to' Guide for Instructors, *Astronomy Education Review*, 1, 1. http://aer.noao.edu/AERArticle.php?issue=1&section=1&article=1.

Callister, J. C., Higham, W. J., Mayer, V. J., Sproull, J. C., & Stroud, S. 1988, *AGI/NSTA Earth Science Examination*, Washington, D.C.: National Science Teachers Association.

Campbell, D. T., & Stanley, J. C. 1963, *Experimental and Quasi-Experimental Designs for Research*, Boston: Houghton Mifflin.

DeLaughter, J. E., Stein, S., Stein, C. A., & Bain, K. R. 1998, Preconceptions Abound Among Students in an Introductory Earth Science Course, *EOS Transactions*, 79, 429.

Frechtling, J. 1995, *Footprints: Strategies for Non-Traditional Program Evaluation*, (NSF 95-41), Arlington, VA: National Science Foundation.

Frechtling, J. 2002, *The 2002 User-Friendly Handbook for Project Evaluation*, (NSF 02-057), Arlington, VA: National Science Foundation.

Frechtling, J., & Sharp, L. 1997, *User-Friendly Handbook for Mixed Method Evaluation*, (NSF 97-153), Arlington, VA: National Science Foundation.

Frechtling, J., Stevens, F., Lawrenz, F., & Sharp, L. 1993, *User-Friendly Handbook for Project Evaluation, Science, Mathematics, Engineering and Technology Education*, (NSF 93-152), Arlington, VA: National Science Foundation.

Hannah, J. L. 1996, What NSF Expects in Project Evaluations for Educational Innovations, *Journal of Geoscience Education*, 44, 412.

Hufnagel, B., Slater, T. F., Deming, G., Adams, J. P., Lindell, R., Brick, C., & Zeilik, M. 2000, Pre-Course Results from the Astronomy Diagnostic Test, *Publications of the Astronomical Society of Australia*, 17, 152.

Libarkin, J. C. 2001, Development of an Assessment of Student Conception of the Nature of Science, *Journal of Geoscience Education*, 49, 435.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. 2004, *Evaluation: A Systematic Approach*, 7th ed., Thousand Oaks, CA: Sage Publications.

Space Telescope Science Institute. 2005, *Elements of a Program Evaluation*, http://ideas.stsci.edu/Evaluation.shtml.

Zeilik, M. 2003, Birth of the Astronomy Diagnostic Test: Prototest Evolution, *Astronomy Education Review*, 1, 46. http://aer.noao.edu/AERArticle.php?issue=2&section=2&article=5.