

# American Journal of Evaluation

<http://aje.sagepub.com>

---

## Developing Evaluation Capacity in Extension 4-H Field Faculty: A Framework for Success

Mary E. Arnold

*American Journal of Evaluation* 2006; 27; 257

DOI: 10.1177/1098214006287989

The online version of this article can be found at:  
<http://aje.sagepub.com/cgi/content/abstract/27/2/257>

---

Published by:



<http://www.sagepublications.com>

On behalf of:

American Evaluation Association

**Additional services and information for *American Journal of Evaluation* can be found at:**

**Email Alerts:** <http://aje.sagepub.com/cgi/alerts>

**Subscriptions:** <http://aje.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations** <http://aje.sagepub.com/cgi/content/refs/27/2/257>

# Developing Evaluation Capacity in Extension 4-H Field Faculty

## A Framework for Success

Mary E. Arnold  
*Oregon State University*

**Abstract:** Developing evaluation capacity in organizations is a complex and multifaceted task. This article outlines a framework for building evaluation capacity. The framework is based on four strategic methods for teaching evaluation: (a) using logic models for sound program planning, (b) providing one-on-one help, (c) facilitating small-team collaborative evaluations, and (d) conducting large-scale multisite evaluations. The article also reports the results of using the framework successfully with Extension 4-H field educators. 4-H educators who were trained using this method significantly increased their evaluation knowledge and skill and reported feeling more positively about evaluation. In addition, the results show that the 4-H organization as a whole has developed a more positive evaluation culture. This framework for teaching evaluation has potential value for all organizations interested in developing evaluation capacity in regional educators as well as developing a positive culture of evaluation within the organization.

**Keywords:** *teaching evaluation; evaluation capacity; evaluation culture; organizational development*

Just 5 years ago, most 4-H educators in Oregon knew very little about program evaluation. Five years ago, evaluation practice often involved a hastily-put-together survey to measure something about a program taking place that same evening. Today, 4-H educators are much more likely to seek assistance with conceptualizing an evaluation protocol, developing an evaluation tool, writing a letter of informed consent, or brainstorming ways to disseminate evaluation results to particular audiences. In the past few years, the evaluation capacity of 4-H educators has undergone a tremendous evolution of success. This article presents a framework of evaluation training and support that was used successfully to increase evaluation capacity among 4-H field faculty in Oregon. In addition, the results of a study of the faculty trained with the framework are presented as evidence of the framework's success.

---

Mary E. Arnold, 4-H Youth Development Education, Oregon State University, 105 Ballard Extension Hall, Corvallis, OR 97331; phone: (541) 737-1315; fax: (541) 737-1332; e-mail: mary.arnold@oregonstate.edu.

Author's Note: I would like to acknowledge Dr. Molly Engle, Extension Evaluation Specialist at Oregon State University, for her joint contribution to the first two stages of the framework described in this article. Dr. Engle and I worked in similar positions during the early stages of building evaluation capacity with the educators of the Oregon State University Extension Service. We conducted trainings together and worked to make the use of the logic model a central part of program planning and evaluation. Together, Dr. Engle and I discovered the value of the one-on-one approach to teaching evaluation effectively. Our early work together helped set the stage for the framework that is now in place in the 4-H Youth Development Education Department.

American Journal of Evaluation, Vol. 27 No. 2, June 2006 257-269  
DOI: 10.1177/1098214006287989  
© 2006 American Evaluation Association

## Building Evaluation Capacity

Stockdill, Baizerman, and Compton (2002) emphasized that the process of evaluation capacity building is dependent on the context of the organization, that it happens through intentional and sustained effort, that it requires guidance, and that as a result evaluation becomes ordinary practice within the organization. Furthermore, the authors outlined several factors that must be considered to understand the process of evaluation capacity building efforts. These include the role of stakeholders, the source of the demand for evaluation, the multilayered and often quite nuanced levels of an organization, the types of methods and training needed, and the resources and flexibility required for successful capacity-building efforts.

Contextual understanding of the organization is becoming increasingly important in developing methods to build evaluation capacity. As Torres and Preskill (2001) pointed out, there has been a recent shift from large-scale external evaluations, from which the results were rarely used within the organization, to more flexible, internal evaluations that use results to learn about, understand, and improve local practice. This change requires that organizations develop a culture that understands and appreciates the value of evaluation in order to help ensure evaluation success.

Douglass, Boyd, and Gundermann (2003) explored the development and evaluation culture within organizations, particularly focusing on the role of evaluation capacity in public agencies. The authors emphasized that the skills, knowledge, and attitudes of individuals within the organization are important factors in determining evaluation competence. Furthermore, they proposed that individuals within an organization will fall on a continuum of evaluation capacity that ranges from *doubters* (individuals who see little value in evaluation) to *scholars* (those who develop considerable expertise in evaluation and actively share their expertise outside the organization). Douglass et al. (2003) proposed that the ideal picture of evaluation capacity in public agencies should have the highest number of individuals who are *practitioners* of evaluation (those who have and use the level of evaluation skills expected and needed by the organization), with very few doubters or scholars (Figure 1).

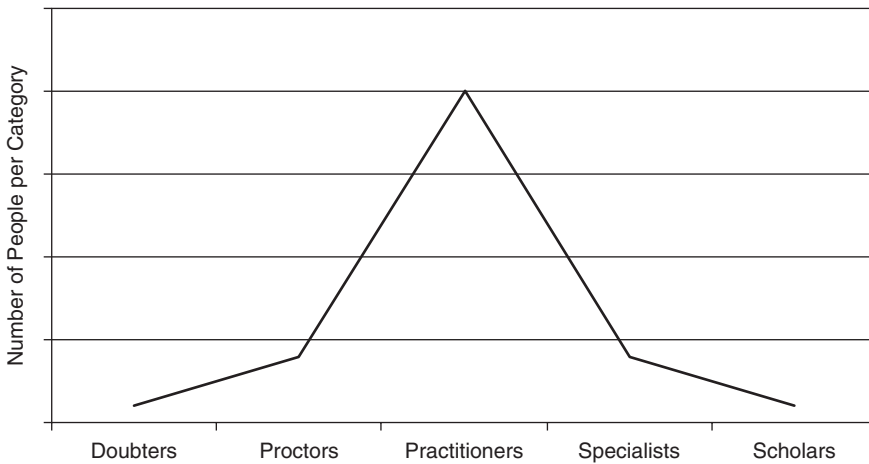
## An Evaluation Capacity Building Training Framework

The four-part framework described in this article (Figure 2) is based on an up-front acknowledgment that 4-H field faculty do not usually have training in evaluation, research methods, or statistics. In most cases, county 4-H educators have degrees in youth development, agriculture education, nutrition, or family and consumer sciences. Any research training that was obtained is often left far behind as the realities and pressures of maintaining an active and vital county 4-H program take over.

The framework is further based on the recognition that maintaining an active county program is not enough to fulfill the job requirements of a county 4-H educator. As members of the faculty, field educators are expected to make scholarly and creative contributions to the field of positive youth development. Conducting formal evaluations of educational programs to understand the effectiveness of program methods and outcomes, as well as sharing the results with others through peer-reviewed presentations at conferences, is one of the main ways that such scholarly contributions are made. In addition, as traditional funding sources dry up, more and more educators are relying on grant funding for their programs; funding that comes with increasing expectations for program evaluation.

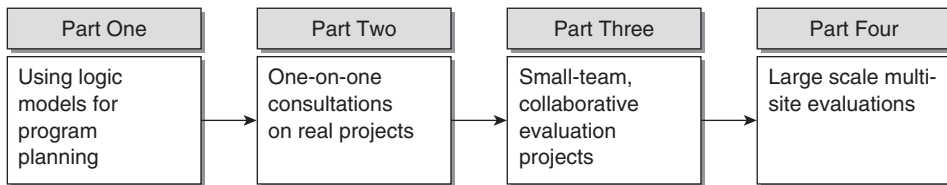
With this in mind, the framework was developed to change the evaluation culture of the 4-H program by increasing evaluation capacity, while at the same time reducing the stress and anxiety that often surround program evaluation. Although there are probably other ways of teaching evaluation and building evaluation capacity, we have found the approach presented here to be effective in meeting the needs of the educators with whom we work. The

**Figure 1**  
**Ideal Evaluation Capacity in Organizations**



Source: Reprinted with permission from M. Douglass, H. Boyd, and D. Gundermann (2003, November). *Nurturing the development of an evaluation culture in public educational agencies*. Paper presented at the annual conference of the American Evaluation Association, Reno, NV.

**Figure 2**  
**Evaluation Capacity Building Framework**



framework works well with individuals who possess a wide variety of evaluation experience. For those just beginning in evaluation, we can start at the beginning. For those with some evaluation experience, we can begin where we need to. The framework also matches the evaluation needs of our organization. For example, we have need for the evaluation of local programming as well as the evaluation of major multisite efforts. This means that every educator must have the basic skills needed to plan and evaluate local programs, but there must still be room for larger scale evaluations. Although local educators do not typically conduct the large-scale evaluations, their role in planning the evaluation and collecting data is critical. Therefore, we need to have an evaluation culture that values evaluation efforts at all levels and that possesses the basic evaluation skills needed at the local level.

The framework includes four components:

1. Providing clear expectations and training in order to increase educators' ability to plan educational programs through the use of logic models. This includes identifying indicators for success and potential points of evaluation at all levels of the model. Using logic models helps educators

- develop capacity for sound program planning and understanding potential evaluation areas and methods.
2. Providing one-on-one consultations for individual evaluation projects rather than general trainings on evaluation. In this way, educators develop evaluation capacity on real projects that hold great personal value.
  3. Conducting program evaluations in small teams, across counties. Doing so allows county educators to work collaboratively to plan and conduct an evaluation, with each team member playing a vital role in the evaluation.
  4. Conducting at least one statewide, multisite evaluation each year. This allows educators to reap the benefits of evaluation without having to do all the work. Statewide evaluations have played a critical role in raising educator interest in the value and usefulness of evaluation.

This framework of training and support for building program evaluation capacity has contributed to changing the culture of evaluation in the Oregon 4-H Program. The framework has helped move the faculty from a body of “doubters” or “proctors” to one of “practitioners” (Douglass et al., 2003). The framework has helped the program accomplish a critical first step in developing evaluation capacity and is paving the way for continued development within the organization.

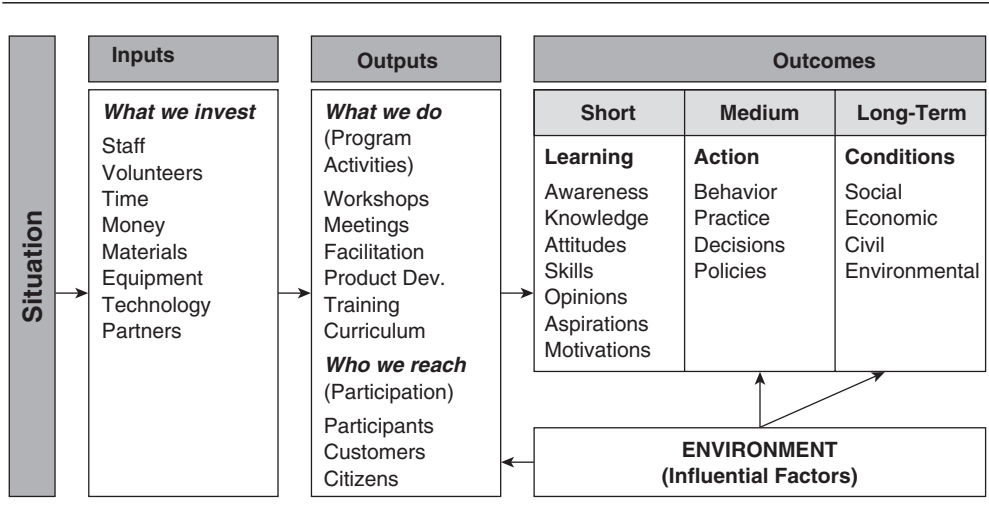
#### *Developing Logic Models as a Common Ground for Program Planning and Evaluation*

Most evaluators will agree that quality program evaluation begins with sound program planning. Yet all of us who teach evaluation know how difficult it can be to convince our clientele of this fundamental principle. The lives of 4-H faculty are packed with meetings, volunteers, youth, and events that encompass everything from a “beef” weigh-in for a county fair to after-school programs to increase science skills. 4-H educators are always on the go and rarely have time to thoroughly plan programs in a way that supports evaluation. As a result, the evaluation is often an afterthought, hastily put together to document *something* about the programs they do. This approach to evaluation was quite common in the 4-H program in Oregon 5 years ago. As we considered how to move forward, we decided to begin with developing skills for sound program planning through the use of logic models, while emphasizing how quality evaluation grows out of good program planning.

The use of logic models (Figure 3) as an aid in program evaluation has received considerable attention in recent years (Brown & Kiernan, 1998; Hatry, van Houten, Plantz, & Greenway, 1996; Hernandez, 2000). Primarily because of the need to understand better the effects and impacts of our programs, and supported by the education outreach efforts of the University of Wisconsin Extension, an awareness of the usefulness of logic modeling in program planning and evaluation has swept Extension services across the country. In a nutshell, a logic model serves as a planning and evaluation tool. As a planning tool, it can help educators identify what they will put into a given program (inputs), what they hope to do, and whom they hope to reach (outputs). The model also identifies short-, medium-, and long-term outcomes for the program. As an evaluation tool, it can help educators identify what and when to evaluate.

Logic modeling can, however, be quite an abstract exercise if not placed in a context that educators can easily grasp and apply. Even when we taught logic modeling skills using familiar and 4-H examples, there often remained a gap between what was being presented and how the educator would actually use it when he or she returned home after the training. The realization that educators were attending trainings and learning about logic models but were not implementing what was learned led us to develop a different, more effective approach.

**Figure 3**  
**Example of a Logic Model**



*Real-Time and Real-Project Consultations*

Despite the success we had in using logic models to frame evaluation, we identified two problems that prevented the successful application of the logic model training we provided. The first problem was that educators had difficulty taking the information about logic modeling from the training and applying it to their own projects. Although educators reported leaving the training with a good understanding of the logic model process, they often reported that it was challenging to apply the model on their own. The second problem was that the timing of logic model trainings was not always conducive to a productive application of the model to individual projects. Educators would report attending training and being excited about using the logic model, only to find that 6 months would slip by before they were ready to apply the things learned in the training. This lapse of time between learning and application often resulted in their abandoning the effort all together.

What emerged as a result of these problems was a one-on-one “real-time” consultation method of teaching evaluation (Engle & Arnold, 2002). Although somewhat time-consuming, we found that educators could more easily grasp the concepts of program planning and evaluation when they could be applied immediately to a real project in which the educator was invested. The one-on-one consultations are driven primarily by the educator, who requests help with his or her evaluation project. Beginning with the common ground of a logic model program plan, the goal of the one-on-one project consultations is to meet the educator where he or she is and develop evaluation capacity from that point forward. This means having to be flexible in our approach and be willing to take small steps in the right direction. In some cases, just getting the educator to articulate learning outcomes for a program is a big success. In other cases, educators can easily identify their program outcomes and need assistance with how to measure them. Other educators now even conduct their own data analysis but turn to us for help in interpretation. Our experience shows, however, that regardless of where an educator begins, he or she always can move forward in developing skills for planning and evaluation. The key, we have found, is to make things simple and provide lots of support and

encouragement. Inevitably, once an educator has some success with evaluation, often through scholarship or recognition, he or she is eager to learn more. It is very rare that we return again and again to the same place with any educator.

### *Collaborative Team Evaluations*

The third component of the framework is to work collaboratively with small groups of educators to evaluate a common program. This differs from the one-on-one consultations in that the project is not usually one person's individual project but rather a team or collaborative effort that may take place over a larger regional area. In addition, those who participate in the collaborative team evaluations have typically already received one-on-one assistance with smaller projects. We saw the movement from one-on-one projects toward larger collaborative projects as evidence of increased interest and skill with evaluation.

Working in small, collaborative teams allows educators to experience all phases of the evaluation process without having to "go it alone." This includes collectively articulating what to evaluate, establishing evaluation questions, developing an evaluation protocol, conducting the evaluation, and interpreting and disseminating results. In collaborative projects, the evaluation trainer can provide assistance with more than one person at a time while maintaining the "real-time, real-project" approach. Collaborative evaluations allow opportunity for exchanges of ideas and methods and can diminish feelings of uncertainty or inadequacy as educators learn and practice evaluation skills together.

A wonderful example of a small collaboration occurred in the spring of 2004 when an educator contacted us with a very general desire to "evaluate the impact of her 4-H program." This educator was one of the early adopters of evaluation and had conducted several small evaluations on her own. She now desired to do something bigger and more comprehensive. We suggested that she invite a few 4-H educators in neighboring counties to come together for a meeting to discuss a collaborative evaluation. At the meeting, the trainer played a mentoring/facilitator role, teasing out program nuances and potential evaluation questions. Very quickly, the group identified programming for older youth as an interesting and unique educational program they all provided. Together, we formed the evaluation questions about the impact of older youth programming, discussed methodology, analysis, and a protocol. We divided up the tasks to be done, and although the trainer agreed to oversee the evaluation as principal investigator, each educator had an important and vital role to play in conducting the evaluation. Throughout the process, we focused on empowering the educators to see our work together as "their evaluation" and the results of "their programs" and to build evaluation skills at each stage of the project.

The results of this collaboration were inspiring. Through the project, we collected some of the first data about the impact of older youth programming on youth developmental outcomes. In addition, the collaborative approach with a small group of educators allowed room for everyone to have an important role and to experience ownership of the project and its results. As a result of participating in the collaborative, each educator gained skills and confidence in evaluation. In addition, the nature of the research questions invited the use of ANOVA and regression analysis, techniques to which the educators had never been exposed. Because of this, the educators learned about more advanced analysis, which in turn increased their interest for exploring even more complex evaluation questions and designs.

### *Multisite Evaluations*

Over the years, the 4-H Youth Development program has struggled with the question of how to evaluate the success of its programs. One of the main factors involved in this struggle

is that programs vary from county to county and from state to state (Meyers, 1980). Although individual local programs are often very successful, documenting that success in a rigorous and systemic way can be difficult, partly because local programs rarely have the number of participants needed to make any generalized statement about the program. One way to address this issue is to conduct multisite evaluations, which is the process of conducting evaluations of the same program that is taking place at different geographic locations (Straw & Herrell, 2002).

Multisite evaluations make up the fourth component of the framework and provide an excellent way to conduct large-scale evaluations that benefit local educators. Each year, we select a 4-H program that is offered statewide and conduct a multisite evaluation. In the past several years, we have conducted statewide evaluations on new 4-H Leader training programs, youth leadership training programs, and 4-H residential camp programs. Unlike the collaborative team evaluation, individual educators have only small roles in the evaluation, usually helping with data collection. Even with minimal participation on the educator's part, statewide evaluations play an important role in raising awareness of evaluation and the results evaluation can produce. Educator participation in these evaluations is voluntary, but as the benefit to educators became clear, participation increased. In the summer of 2004, we conducted a statewide evaluation of residential 4-H camps, with more than 850 campers and 200 older youth counselors participating from 17 camp sites across the state. The results of the study are far more robust and representative of 4-H camping in Oregon than any individual county camp evaluation would have produced.

Several key factors appear to make multisite evaluations work. First, we chose programs with common goals and methods across sites. This is important for obtaining results with any validity. Using logic models to identify common program goals and outcomes is very useful. The second factor is having buy-in from the county educators who participate. At the end of the evaluation, each educator receives a detailed evaluation report containing results pertinent to his or her county. This information is highly valuable to county programs and stakeholders and thus serves as a motivator for participation. Finally, having a training session for all people involved in the evaluation prior to embarking on the project is important. The training session allows for discussion about the philosophy of the evaluation, its goals and methods, and allows time for detailed question-and-answer time, so that all participants leave with a clear understanding of what is to take place.

## Evaluation of the Training Framework

### Method

Deliberate efforts to build 4-H educator evaluation capacity began in the fall of 2000 with the introduction of logic model trainings. The complete framework was developed and implemented over a 5-year span between 2001 and 2004. By the fall of 2004, we began to suspect that the evaluation culture of the 4-H program had changed considerably, so we embarked on a study to determine the changes that had taken place.

#### *Participants*

Thirty-seven field educators who were identified as the main 4-H faculty field educators (i.e., considered the 4-H "agents" in the county) were invited to participate in an online survey. Recent evidence suggests that online surveys are an effective and efficient way to gather evaluation data (Kiernan, Kiernan, Oyler, & Gilles, 2005). Responses were received



from 30 participants, for an 81% return rate. Eighty percent of the respondents ( $n = 24$ ) had been 4-H agents since 2000 or earlier, and the rest joined the organization in 2001 or later.

### *Procedures*

Participants were contacted via an e-mail inviting them to take part in the study. The questionnaire was constructed using a professional online survey company (Survey Monkey). Participants followed a link in the e-mail invitation to the survey site. A follow-up e-mail was sent after 2 weeks and again after 4 weeks to those who had not yet responded to the questionnaire.

### *Instruments*

A questionnaire was developed specifically for the study. Face validity for the instrument was established through peer review by colleagues familiar with the goals of the training framework. No attempt was made to establish reliability for the instrument. The first seven items related to changes in evaluation capacity, including overall evaluation capacity, knowledge of logic modeling, skill in logic modeling, knowledge in program evaluation, skill in program evaluation, attitude about program evaluation, and competence in program evaluation.

Using a retrospective pretest methodology (Pratt, McGuigan, & Katzev, 2000), participants were asked to rate each question based on their perceived knowledge and ability today and their perceived knowledge and ability 4 years ago. This type of assessment is commonly referred to as the *perceived-change* method and is used to measure how much program participants feel they have changed as a result of a program (Lam & Bengo, 2003). Concerns have been raised about the potential for program results to be overestimated when measured using a retrospective pretest method, but there is evidence to suggest that this method is useful when the goal of the evaluation is to assess individual perceptions of change, as we did in this study (Hill & Betz, 2005).

A second set of seven questions asked participants to rate their agreement with how much the training program helped them to develop skill in the areas of articulating program outcomes, planning educational programs, conducting a learning assessment, preparing an application for the Institutional Review Board for the use of human subjects, developing evaluation tools, interpreting evaluation results, and writing up evaluation findings. Participants rated each item on a scale from 1 (*strongly disagree*) to 5 (*strongly agree*).

The scale of evaluation levels in an organization proposed by Douglass et al. (2003) identifies five levels of evaluation practice: (a) doubters, (b) proctors, (c) practitioners, (d) specialists, and (e) scholars. Using this scale, participants were asked to rate their placement on the scale 4 years ago and now.

Last, participants were asked to rate how valuable the training framework is to their work, to the 4-H program, and to their academic and scholarly work. Participants rated each item on a scale from 1 (*not valuable at all*) to 5 (*extremely valuable*).

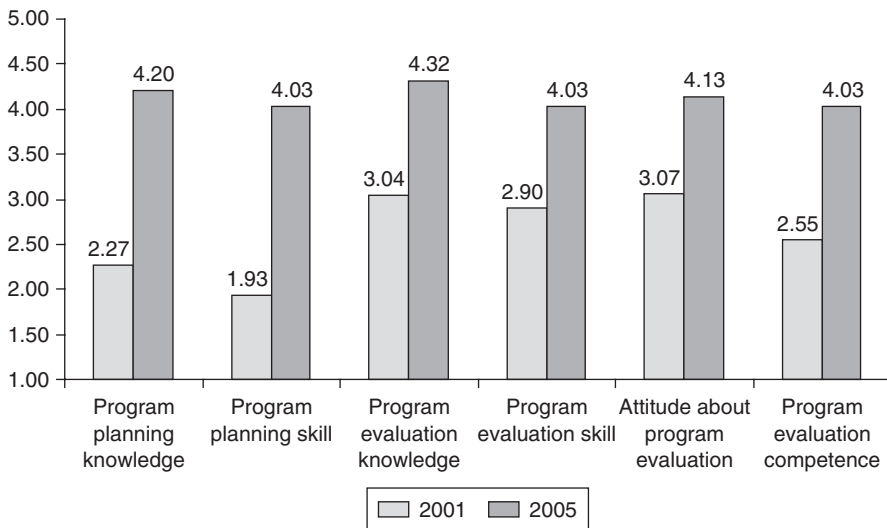
In addition to the quantitative data gathered via the survey, respondents were invited to provide narrative data to illustrate how the training framework had affected the way they thought about and practiced program evaluation. These data were used to support and illustrate the conclusions drawn from the study.

**Table 1**  
**Paired *t* Tests for Changes in Evaluation Knowledge and Skill**

	<i>N</i>	<i>M</i> Before	<i>M</i> After	Mean Difference	<i>SEM</i>	<i>t</i>	Significance <i>df</i>	Two-Tailed
Program planning knowledge	30	2.27	4.20	-1.93	0.26	-7.50	29	.00
Program planning skill	30	1.93	4.03	-2.10	0.21	-9.96	29	.00
Program evaluation knowledge	8	3.04	4.32	-1.29	0.22	-5.92	27	.00
Program evaluation skill	29	2.90	4.03	-1.14	0.19	-5.98	28	.00
Attitude about program evaluation	30	3.07	4.13	-1.07	0.22	-4.75	29	.00
Program evaluation competence	29	2.55	4.03	-1.48	0.23	-6.43	28	.00

Note: *SEM* = standard mean error.

**Figure 4**  
**Mean Ratings for Changes from Pre to Post**



## Results

The seven retrospective pretest questions related to changes in evaluation knowledge and skill were analyzed using a paired *t* test. Significant changes were found for each item. Table 1 presents the results of the paired *t* tests; Figure 4 presents a graphic of the mean changes.

Mean ratings for the questions related to how much the training program helped participants develop evaluation skills ranged from a high of 4.10 (articulating program outcomes) to a low of 3.23 (writing up evaluation results). Table 2 presents the range, mean ratings, and

**Table 2**  
**Range and Mean Ratings for Knowledge and Skill Development**

	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>
Articulating program outcomes	30	2	5	4.10	0.76
Planning educational programs	30	2	5	4.00	0.69
Conducting a learning assessment	29	2	5	3.90	0.67
Developing evaluation tools	30	2	5	3.80	0.76
Preparing an IRB application	30	2	5	3.67	1.15
Interpreting evaluation results	30	2	5	3.50	0.90
Writing research reports	30	2	5	3.23	1.01

Note: IRB = Institutional Review Board.

**Table 3**  
**Range and Mean Ratings for Value of Training Program**

	<i>N</i>	Minimum	Maximum	<i>M</i>	<i>SD</i>
Value to 4-H academic and scholarly work	30	3	5	4.70	0.53
Value to the 4-H program	30	3	5	4.67	0.55
Value to the faculty member's work	30	2	5	4.40	0.77

standard deviations for each evaluation skill item. Mean ratings for the value of the training program ranged from a high of 4.70 (value to academic and scholarly work) to a low of 4.40 (value to the faculty member's work). Table 3 presents the range, mean ratings, and standard deviations for each of the evaluation value items.

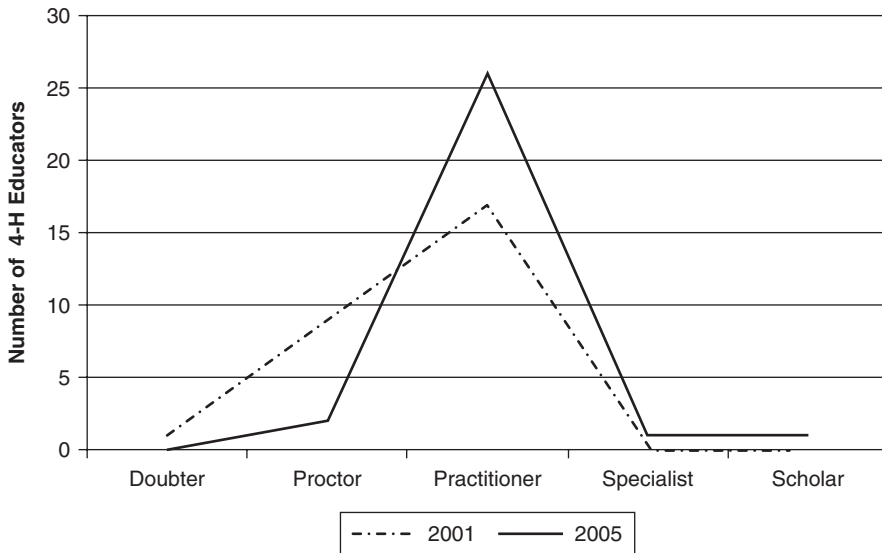
Finally, an analysis was conducted of the participants' ratings of their level of evaluation practice now and 4 years ago. Frequencies for each level and each time were plotted on a line graph (Figure 5). The graph reveals that there has been a distinct movement toward the ideal evaluation culture as proposed by Douglass et al. (2003).

## Conclusion

The four-component framework for building evaluation capacity presented in this article has produced significant results in a very short time. Five years ago, the idea of evaluation was viewed negatively as a "top-down" request. As one educator put it, "Honestly, when this first came about [program planning and evaluation], I did not fully understand the importance. But I went through the motions of providing what was necessary because it was required." Although mandated from administrators, there was little understanding of what the mandate meant except that performance evaluations were linked to evaluative evidence about the impact of the educator's work. This set the stage for a great deal of anxiety and frustration about program evaluation.

The evaluation training framework provided a useful and meaningful way to empower 4-H educators in program evaluation. We began by establishing a common ground through using logic modeling to plan for program evaluations. Although educators are continually developing and refining their logic modeling skills, there is now a common language among 4-H educators. For example, the terms *inputs*, *outputs*, *outcomes*, *indicators*, and *evaluation methods* are part of our organization's everyday language. As one 4-H educator said,

**Figure 5**  
**Frequency of Ratings of Evaluation Level in 2001 and 2005**



My understanding and use of logic model terminology has definitely changed from 5 years ago when my knowledge in this area was practically nonexistent. Because of the training program in this area, the logic model is now a natural part of my process in program planning and evaluation.

Another educator expressed how he uses the logic model:

Whether on paper or as mental exercise, I have used the logic model to help myself and others better understand state and local 4-H programs. The logic model has played an integral role in how I design programs, develop evaluation instruments, and critically evaluate programs. Additionally, I use the logic model as a precursor to grants and other fund-raising initiatives with success.

The one-on-one, “real-time” evaluation assistance helped educators gain skill and confidence in program evaluation. As confidence and skills grow, so does the educator’s desire to learn more about his or her programs. It is at this point that we often witness evaluation taking on an internal value for the educator. As one educator reported, “In working together one-on-one on two different program evaluation projects, I gleaned knowledge and skills that gave me the confidence to take on evaluation development on my own.” Another reported how his confidence had developed in very specific ways, such as preparing evaluations for county, regional, and statewide 4-H programs, interpreting results for himself and others, and taking the lead on designing evaluations of programs done in collaboration with other organizations.

The small collaborative evaluation projects add to the basic skills gained through individual program evaluations. The evaluations we have done in this way have allowed educators to learn new skills and gain experience by asking more complex questions and using more

sophisticated analysis techniques. For many, this has also changed educators' motivation to conduct evaluations from an external mandate to an internal desire to understand the impact of their programs. As one educator stated,

Although I still have much to learn, I am more aware of the importance of evaluating my programs. Before, I primarily focused on the everyday aspects of delivering programs to youth. Although I have always believed that 4-H makes a difference in the lives of youth, I did not fully understand the importance of learning *how* this program impacts their future. I am definitely more interested in, and aware of, short-, medium-, and long-term outcomes and the importance of including this aspect into our programs.

Finally, statewide, multisite evaluations have played a major role in building evaluation capacity in two main ways. First, educators reap the benefit of the results of the program evaluation while playing only a minor role in the process. This can be an effective way to pique an educator's interest in program evaluation, causing him or her to want to duplicate the evaluation on a local program. Second, as the educator discovers the utility of the evaluation results for describing and promoting his or her work, there is often an increased interest in gaining evaluation skills. As one educator illustrated, "There is a limited arsenal of tactics that you can use in a small, county-only evaluation. Being involved in statewide studies has helped me learn additional tools. Plus, you get enough data to show impact, which is very rewarding."

As noted earlier, most 4-H educators in Oregon are members of the faculty at Oregon State University. With this appointment comes the requirement for scholarship and other creative activity. Developing program evaluation capacity in Oregon's 4-H educators has helped contribute to the educators' scholarly success, primarily because educators now have evidence of effective program methods and outcomes to share with others. One of the main ways that 4-H educators disseminate the results of their work is through peer-reviewed presentations and seminars at professional meetings. In recent years, conferences such as the National Association of Extension 4-H Agents (NAE4-HA) have added "evidence of program success" to the review criteria of conference proposals. Oregon 4-H educators have consistently been selected to present their work at conferences, in large part because of their ability to measure the outcomes and impacts of their programs.

It is important to state that the success of this framework has been dependent on several key factors, and we would be remiss if we did not mention the infrastructure behind the framework. First, the 4-H program in Oregon recognized the need to have "in-house" expertise in order to build evaluation capacity. As such, there is a full-time 4-H specialist dedicated to providing program evaluation training and support to county educators. This person is responsible for ensuring that program evaluations take place and that educators receive the support they need to build skill and conduct their own evaluations. This framework describes the method through which the evaluation specialist has learned to do his or her job effectively. Second, the 4-H department has provided needed resources, such as computers, software, money, space, and professional development support to allow the evaluation specialist to provide training and support to county educators. And finally, the department's position on the campus of Oregon State University has provided access to graduate students with current and sophisticated abilities in research methods and data analysis. The graduate students play a key role through their expertise in data entry, analysis, and interpretation.

It is clear to us that the success of this framework is dependent on a supportive infrastructure. Nonetheless, we believe this framework provides the blueprint of an effective strategy for building evaluation capacity among groups of educators who may not have education or experience in evaluation. We think the framework can be particularly effective in organizations that have

educators delivering similar programs in separate locations. The framework provides a way to increase evaluation capacity in a way that is supportive and unifying.

## References

- Brown, J. L., & Kiernan, N. E. (1998). A model for integrating program development and evaluation. *Journal of Extension*, 36(3). Retrieved October 27, 2004, from <http://www.joe.org/joe/1998june/rb5.html>
- Douglah, M., Boyd, H., & Gundermann, D. (2003, November). *Nurturing the development of an evaluation culture in public educational agencies*. Paper presented at the annual conference of the American Evaluation Association, Reno, NV.
- Engle, M., & Arnold, M. E. (2002). When the student is ready: Capitalizing on an opportunity to teach evaluation. Unpublished manuscript.
- Hatry, H., van Houten, T., Plantz, M. C., & Greenway, M. T. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: United Way of America.
- Hernandez, M. (2000). Using logic models and program theory to build outcomes accountability. *Education and Treatment of Children*, 23(1), 24-40.
- Hill, L. G., & Betz, D. L. (2005). Revisiting the retrospective pretest. *American Journal of Evaluation*, 26, 501-517.
- Kiernan, N. E., Kiernan, M., Oyler, M. A., & Gilles, C. (2005). Is a Web survey as effective as a mail survey: A field experiment among computer users. *American Journal of Evaluation*, 26, 245-252.
- Lam, T. C. M., & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation*, 24, 65-80.
- Meyers, J. M. (1980). Integrity versus pragmatism: Planning the national 4-H evaluation. *Education and Urban Society*, 13(1), 13-36.
- Pratt, C. C., McGuigan, W. M., & Katzev, A. R. (2000). Measuring program outcomes: Using retrospective methodology. *American Journal of Evaluation*, 21(3), 341-349.
- Stockdill, S. H., Baizerman, M., & Compton, D. W. (2002). Toward a definition of the ECB process: A conversation with the ECB literature. *New Directions for Evaluation*, 93, 7-25.
- Straw, R. B., & Herrell, J. M. (2002). A framework for understanding and improving multi-site evaluations. *New Directions for Evaluation*, 94, 5-15.
- Torres, R. T., & Preskill, H. (2001). Evaluation and organizational learning: Past, present, and future. *American Journal of Evaluation*, 22, 387-395.