GUIDE FOR PROGRAM

EVALUATION

by

Maud Roucan-Kane, Joseph Suttles

Working Paper #09-

February 2009

Dept. of Agricultural Economics

Purdue University

GUIDE FOR PROGRAM EVALUATION

by
Maud Roucan-Kane, Joseph Suttles
Dept. of Agricultural Economics, Purdue University
West Lafayette, Indiana 47907-2056
mroucan@purdue.edu; jdsuttle@purdue.edu
Working Paper #09-05
February 2009

Abstract

Extension programs must be evaluated in order to assess their value to participants, institutions, funders, and all other stakeholders. Evaluations can be especially useful when the program has specific objectives that are measurable, either qualitatively and quantitatively.   There are many different methods and formats for evaluation, and choosing the correct evaluation can be critical to ensuring the accuracy and relevance of the evaluation results. Considerable thought must be put into determining the correct method for an evaluation, and they must always be focused on the specific objectives of the extension program. After administering an evaluation, the results should be communicated back to the stakeholders of the program in an effective manner and, after some deliberation, the program should be considered for alteration if deemed necessary. This paper aims to be a 'how-to' guide for development, administration, and appraisal of evaluations and evaluation results for a broad spectrum of extension programs.

Keywords: extension program, evaluation, LOGIC

JEL Codes: Q16, I21, I23

GUIDE FOR PROGRAM EVALUATION

Maud Roucan-Kane, Joe Suttles

Center for Food and Agricultural Business, Department of Agricultural Economics, Purdue University

Throughout their long history, extension and continuing education programs have proven to be a useful tool for institutions to reach-out to their surrounding communities and respective fields by bringing cutting-edge research easily within grasp of those who would most benefit from it. Over 100 years ago, the United States Congress formed the Extension System with the explicit purpose of improving production practices in agriculture, and has since maintained the Cooperative Extension Services role at doing just that (see www.csrees.usda.gov/qlinks/extension.html). Even today, in the age of nearly instant and limitless information, extension is still a leading avenue for farmers to learn and improve their farm operations – for their own betterment and the betterment of the nation as a whole. However, developing and presenting educational programs is not free. All stakeholders, particularly program sponsors, have an interest in ensuring that programs are delivering a value to participants that justifies the cost of delivering a program.

One of the best ways to determine if a program is worthwhile is through a simple evaluation. That's pretty straightforward, but by what methods and under which criteria should it be scrutinized? Should all programs be evaluated, or only those for which it can be done cost effectively? The following sections will highlight some of these issues and be a guide for formulating and implementing an extension program evaluation procedure.

What is the purpose of an evaluation?

Before making any determinations about when or how to evaluate a program, it's important to understand exactly what the purpose of an evaluation is. There are a variety of uses for evaluations.

The most straightforward use for an evaluation is to serve as an identifier of the strengths and weaknesses of a program. The information can be used to improve the same program in future settings, or to help in development of similar programs. The information from a well-structured evaluation can give very specific and constructive criticisms that can help answer questions like: Is there any subject matter that should be added? Which teaching styles worked? Which instructors should be retained?

An evaluation may even yield information about the usefulness of the program as a whole. After comparing across programs, it may be clear that an extension program should be dropped altogether. Or, through time, evaluations can prove useful as a gauge for the progression of a program.

Evaluations can also serve as a good public relations tool. Filling out an evaluation at the end of a program helps the participants feel that program coordinators and instructors are truly

concerned about what they think. Participants who walk away feeling just a little better about a program because they feel that program administrators were sincerely concerned with their opinions may be a boon to word-of-mouth advertising as the participants discuss the program with their friends and family. Also within the public relations scope, evaluations can be part of a more formal advertising campaign for a program. It likely means much more to a potential participant that previous participants have rated a program well or to hear a participant's testimonial rather than simply reading a brochure touting the supposed benefits.

When should a program be evaluated?

The use of evaluations is not always justified. Administering an evaluation has a cost, both in time and materials, so it only makes good sense that the expected benefits from an evaluation outweigh the costs of the process. Estimating the benefits of an evaluation is a subjective activity, best handled on a case-by-case basis. However, there are some general guidelines that are useful to follow when trying to determine the need for an evaluation.

One situation where evaluations may be a bad idea would be for a program that is expected to have negative feedback. Negative feedback can often serve as a great learning tool for program managers/presenters, but, for instance, if the sponsors of a particular program would not deal with negative criticism well, it may be best to avoid the evaluation process altogether. After all, if program administers are expecting negative feedback to begin with, perhaps they can communicate the program's shortcomings in a more subtle and tactful way than with terrible evaluation results.

As a general guide, the following characteristics of an extension program make it a good candidate for evaluation:

- The program contains multiple objectives that are measurable
- The stakeholders expect an evaluation of the results
- There are opportunities and sufficient resources available to improve a program if necessary
- There would be a willingness and desire to improve the program if deemed helpful
- A similar program is expected to be offered in the future which may be improved through use of evaluation results

This list is certainly not exhaustive. The format, goals, and circumstances of every program are different, making it impossible to create a simple "evaluate" or "don't evaluate" criteria, but the above guidelines should be considered in determining the need for a program evaluation.

Which Evaluation Method

Once it has been determined that a program warrants evaluation, the next step is to answer the question; what evaluation method should be used? This is not a straightforward question to

answer because, once again, the determination will need to be made on a case-by-case basis. The evaluation method(s) to use will depend on the audience, the type of information desired, the availability of resources, and the quantity of information that will be useful. Not every topic and issue can be evaluated, and there is a trade-off between quantity and quality.

There are a multitude of methods and formats for evaluation. Each has its own strengths and weaknesses, and, if sufficient resources are available, it may be wise to use more than one in combination to overcome the weaknesses of any one given method. The following is a non-exhaustive list of the different types of evaluations with a brief explanation of each. Also, please refer to Table 1 for a concise comparison of the strengths and weaknesses of the different evaluation types.

- Questionnaires/Surveys/Checklists contain a set of questions/items addressed to someone for comments. They are used to collect information from many respondents without direct contact between the evaluators and the participants.
- Interviews consist of a discussion between two parties to exchange information about a program. They are more personal and are used to gather information in person or by telephone.
- Focus groups/Open forums/Verbal feedback sessions gather opinions about a program through open discussion. They allow for in-depth comments with a small group selected from a larger group of participants.
- Observation examines gestures, facial comments, activities, etc to determine one's opinions about the program.
- Knowledge assessments assess participants' improvement in knowledge, as a result of the program, through the use of pre-tests and post-tests, the creation of presentation, the write-up of reports, etc.
- Performance assessments evaluate participants' improvement in skills, as a result of the program.
- Listening posts/Informal conversation help gather a lot of feedbacks from a program by just listening and taking mental notes.
- Box of ideas, comments, suggestions can also be a way for participants to anonymously provide short comments.
- Self evaluation consists of an evaluation of the program by those conducting the program (instructor, extension educator, program/event manager).
- Testimonials are written or oral comments of a participant's program experience.

Table 1. Strengths and Weaknesses of Common Evaluation Tools

Evaluation Type	Strengths	Weaknesses
Questionnaires/Surveys/Checklists	- Easy and inexpensive to administer - Many possible formats: online, on paper, fax, etc - Lots of quantitative data - Easy to summarize	- Specific feedbacks are limited - Some of the participants' comments may be hard to interpret - Comments are limited to the topics listed unless an additional comment section is added

Interviews (phone, in person, and more)	• More qualitative data • Easy to clarify an answer	• Potential interviewer bias • Respondents may not be truthful unless the interview is done by a third party
Focus group/Open forum/Verbal feedback session	• More qualitative data • Easy to clarify an answer • The group format generates more ideas • Participants react to others' ideas which allows for generalization	• Potential moderator bias • Are the respondents representative of all the participants? • Some participants may dominate the discussion
Observation	• Inexpensive • "Honest" reaction if the participant does not know he/she is being observed	• Observer bias • Wrong interpretation • Is the observed moment representative?
Knowledge assessment	• Many possible formats (tests, presentation, reports, etc) • Easy and inexpensive to administer (at least the tests) • A method to collect outcomes of a program and not opinions	• Test anxiety • A test may affect the overall program experience negatively • Rater bias with some formats
Performance assessment	• Many possible formats (peers' assessment, supervisor evaluation, etc) • A method to collect outcomes of a program and not opinions	• Anxiety • May affect the overall program experience negatively • Rater bias
Listening posts/Informal conversation	• Inexpensive • "Honest" reaction if the participant does not know his/her opinions are being "collected"	• Are the conversations listened to, representative of all the conversations • Some participants may dominate the conversation • May not relate to the topics that have to be measured
Box of ideas, comments, suggestions	• Inexpensive	• Usually few comments will be gathered and may not relate to the topics that have to be measured • Some of the participants' comments may be hard to interpret
Self evaluation	• Inexpensive	• Evaluator bias
Testimonials	• Can be used for marketing purpose effectively • Limited data	• Respondents may not be truthful unless the interview is done by a third party • Comments may not relate to the topics that have to be measured • Some of the participants' comments may be hard to interpret

Further information concerning these methods is available at http://www.managementhelp.org/evaluatn/fnl_eval.htm.
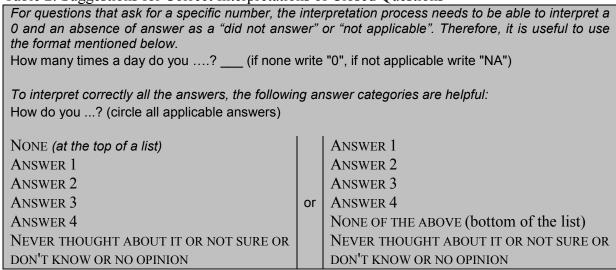
Create the Evaluation Material

Creating the evaluation requires determining the questions of interest and deciding on the answer categories of those questions. The format of the evaluation form is also important.

Examples of Questions and Answer Categories
To avoid bias in the results, evaluations start with some general questions concerning the overall feeling of the program (program focus, format, relevance). The program dose (depth, frequency, amount/quantity) is also of interest. Several evaluations also include logistics and marketing questions. Evaluations sometimes rate instructors and topics. Some demographic information may also be gathered. It is also important to know what learners think about the delivery method and if they have suggestions for possible improvements (Would there be a more effective way? How would you modify the program in the future?). Finally, questions should also address what the stakeholders want to know and what the objectives of the program (see LOGIC model) are (2).

Often, evaluations contain a mix of closed and open-ended questions. Closed questions may ask for a specific answer or may list suggested answers to choose from (see table 2). Suggested answers should be typed in bold and small caps and include at least four categories. Furthermore, repeating the same set of answer categories (when possible) helps the respondent unless the set is repeated too often (2).

Table 2: Suggestions for Correct Interpretations of Closed Questions

For questions that ask for a specific number, the interpretation process needs to be able to interpret a 0 and an absence of answer as a "did not answer" or "not applicable". Therefore, it is useful to use the format mentioned below.
How many times a day do you ....? ___ (if none write "0", if not applicable write "NA")
To interpret correctly all the answers, the following answer categories are helpful:
How do you ...? (circle all applicable answers)

NONE *(at the top of a list)*		ANSWER 1
ANSWER 1		ANSWER 2
ANSWER 2		ANSWER 3
ANSWER 3	or	ANSWER 4
ANSWER 4		NONE OF THE ABOVE (bottom of the list)
NEVER THOUGHT ABOUT IT OR NOT SURE OR DON'T KNOW OR NO OPINION		NEVER THOUGHT ABOUT IT OR NOT SURE OR DON'T KNOW OR NO OPINION

Scales may also be used as answer categories for rating questions. Scales can have a number or word basis. Both have advantages and disadvantages (see table 3) and therefore, some evaluations use scale numbers with a word interpretation (2).

Table 3: Advantages and Disadvantages of Scales with Numbers and Words

	Scales with numbers	Scales with words
Advantages	●Appearance of objectivity ●Clear difference between the different categories ●No coding required	●Meaningful ●Words are prevalent as means of communication
Disadvantages	●Numbers are impersonal ●May lead to lack of seriousness and therefore validity ●Numbers are not always meaningful	●Not easy to find differentiated set of words if lots of answer categories are needed ●Require coding

Ranking questions should be avoided particularly when several items need to be ranked. Instead, the participant should be asked to determine the main issue or the three main issues in the list. (2).

Some programs may have pre[1], post and post-post program evaluation (on a stepscale[2] or regular form). Some studies have shown that participants can be more truthful about their pre-program behavior or knowledge if asked after the program. Because participants do not behave consistently, behavioral questions should ask for the frequency, include the word "ever" or propose answer categories not limited to yes or no (see Table 4) (2).

Table 4: Examples of Behavioral Questions

The following type of question can be used as pre or post-behavioral question:
● How often do you …? (circle the most appropriate answer)
    1. almost all the time
    2. frequently
    3. occasionally
    4. hardly at all
    5. not at all
    6. not applicable.

The following type of question can be used as pre or post-behavioral question:
● Do you ….? (circle the most appropriate answer)
    1. yes (if yes, how often in the last 6 months?_____)
    2. not yet
    3. not at this time
    4. plan to do within 3 months

---

[1] It is critical to assess what was done by the participants before the program.
[2] The following url is a link to a tipsheet created by Pennsylvania University on stepscale questions: http://www.extension.psu.edu/evaluation/pdf/TS47.pdf or http://www.extension.psu.edu/evaluation/titles.html (tipsheet 47)

> *The following type of question can be used as a post-behavioral question to assess the degree of achievements:*
>
> ● Since the program has ended, to what extent have you been able to do each of the following? (circle the most appropriate answer)
>
> Item 1: Thought about or discussed   Started writing   Completed writing   Not get started   Did before program
>
> or
>
> Item 1: Did before program   Started since program   Plan to do within 3 months   No plans   Not applicable
>
> *Since some people may not have had the opportunity to make any changes yet since the end of the program, an open-ended question may be considered:*
>
> ● Many people may not have had an opportunity to make any changes since the program. If you are one of these, please indicate why?

Questions regarding age, race and marital status are extremely challenging. Age is a sensitive question and hence, should be avoided unless the information is needed. The question should be placed at the end of the questionnaire unless the screening is on age. Age categories should be used (including one higher than the oldest persons of the audience), be relevant to the audience and mutually exclusive (e.g., 20 to 34 years, 35 to 54 years and so on[3]). If the exact age is needed, the Census Bureau has found that asking for the year of birth leads to more accurate answers. Tables 5 and 6 present some recommended race and marital status questions (2).

Table 5: Race Question

> *Many persons identify themselves as belonging to two or more races. Therefore, the use of the US Census Bureau method (reported below) is recommended for questions regarding race.*
>
> How would you describe yourself? (circle all that apply)
> 1. Black/African-American
> 2. Hispanic
> 3. White (Non-Hispanic)
> 4. Other (please specify):

Table 6: Question concerning Marital Status

> *The following question may be used to give people the opportunity to report that they live with another adult besides a legal spouse.*

---

[3] This concept is also valid for income, frequency of behaviors and more.

> What is your marital status? (Circle the appropriate answer)
> 1. Single
> 2. Living with significant other or partner
> 3. Currently married
> 4. Divorced or separated
> 5. Widowed

Open-ended questions are hard to analyze for quantitative research but are great tools and are less subject to interpretation for qualitative research. They may want to be avoided if many different answers are likely to be given by the audience. Instead, a close-ended question with suggested answers should be used. To determine the suggested answers, an open-ended question can be asked the first few times the program is given or as a pre-test given to a few persons. An "other" category should also be added to the suggested answer to give the participant the opportunity to add something not included in the list[4] (2).

Format
Survey research demonstrates that design is more important than length to motivate completion. It has to look easy to do and be consistent. Tables are not recommended. They tend to make the evaluation look hard and if the answer categories are not repeated for each statement, participants' mistakes are likely. Instead, the statements and answer categories should just be listed. For open-ended answers, lines should be shortened and centered to look shorter. Suggested answers should be placed each below the others to give an impression of space (2).

Administer the Evaluation
Before administering the evaluation, it is recommended to pretest the evaluation with a small sample. When the evaluation is administered or advertised, it is necessary to give a rationale[5] (i.e., explain the purpose and the reasons) for the evaluation and show that participation is respected[6] (2).

After the Evaluation
For questions involving numbers, results should be reported on a percentage form for each suggested answer (for example: percentage of participants who give a 1). The mean or the average for each question with numbers should also be mentioned. For open-ended answers, comments should be organized into meaningful categories (concerns, suggestions, strengths, weaknesses), the categories should be labeled, and patterns should be identified. If participants are quoted and their names are reported, the participants should have given their consent (1).

---

[4] The following url is a link to a tipsheet created by Pennsylvania University presenting examples of answer catagories: http://www.extension.psu.edu/evaluation/pdf/TS44.pdf or
http://www.extension.psu.edu/evaluation/titles.html (tipsheet 44)
[5] "Our team wishes to plan similar programs in the future and would like to know what you, as participants, thought of the program."
[6] "Answers will be kept confidential, results will be used and reported in a summarized form."

The results should then be used for marketing and sale purpose. Results should also be communicated to the team and instructors involved in the program. This communication will help the different members determine the success, see where to improve and what to duplicate. Positive results are also good motivation tools. Results (if requested or positive) should also be shared with the clients or grant funders. This may help generate more sales or grants. Finally, some follow-up may be organized and some participants may be contacted to get a better understanding of the comments or ask for their opinion on replacement solutions.

Bibliography

(1) McNamara, Carter. "Basic Guide to Program Evaluation." accessed at

http://www.managementhelp.org/evaluatn/fnl_eval.htm on 07/02/07.

(2) PennState, College of Agricultural Sciences, Cooperative Extension and Outreach,

« Evaluation Tipsheets ? » accessed at http://www.extension.psu.edu/evaluation/titles.html on

07/02/07.

Note: More information is available at

http://www.canr.msu.edu/evaluate/AllTextMaterial/ProgEvaRes.html,

http://www.flinders.edu.au/teach/assess/,

http://www.ccl.org/CCLCommerce/publications/landing/evaluating.aspx?CatalogID=Publicatio

ns, and on "*The Targeted Evaluation Process" by Wendy L. Combs and Salvatore V. Falletta.*