

impact evaluation: a challenge for extension

M. F. Smith
A. A. Straughn

*"We cannot overstress the difficulty
of undertaking impact evaluations."*

Rossi and Freeman¹

Extension has never been more concerned than now with impact . . . impact evaluation, that is. As a result of major emphasis being placed on this type of study in the new accountability/evaluation (A/E) four-year system,² each state and most counties will be affected by federal impact studies, state impact studies, or state requirements for county impact studies.

As a result of interacting with a variety of people from different states and different levels of the Extension organization, it has become clear a lack of agreement exists about what's meant by "impact evaluation." And, until there's agreement on a definition, methodology will be uncertain, staff training will be a hit-or-miss proposition, and aggregation of data from site to site and state to state will yield data of questionable validity and utility.

Impact Evaluation Definition

Cooperative Extension's usual procedure is to contact experts in a field when questions arise about that particular area—a tomato expert for information on new tomato varieties; a child development expert for information about the emotional development of youth. Following a similar

M. F. Smith: Assistant Professor and Program Evaluation Specialist, Institute of Food and Agricultural Sciences, Cooperative Extension Service, University of Florida—Gainesville and *A. A. Straughn:* Associate Professor and Director, Program Evaluation and Organizational Development, Institute of Food and Agricultural Sciences, Cooperative Extension Service, University of Florida—Gainesville. Accepted for publication: May, 1983.

procedure for questions about impact evaluation showed much agreement among evaluation experts about a definition for impact evaluation. For example:

1. In a 1980 publication,³ the Evaluation Research Society Standards Committee listed impact evaluation as *one* category of evaluation “aimed at determining program results and effects, especially for the purposes of making major decisions about program continuation, expansion, reduction, and funding.” In a later publication,⁴ note was made of the challenge to attribute types and amounts of impact to the program rather than to other influences.
2. Patton says that impact evaluations gather data on “the direct and indirect program effects on the larger community of which it is a part.”⁵
3. Cronbach indicates that the basic aim of impact evaluation “is to estimate the net effects or net outcomes of an intervention . . . free and clear of the effects of other elements in the situation under evaluation.”⁶
4. Rossi and Freeman state that “impact assessment is directed at establishing, with as much certainty as possible, whether or not an intervention is producing its intended effects.”⁷
5. Sanders uses the term “impact study” to refer to one “designed to provide information about the outcomes of an Extension service or program and whether such outcomes can be attributed to the program or service rather than to some extraneous circumstances.”⁸

The consensus from these authors is that impact evaluation tries to determine the net effects of programs—net, meaning that effects from other aspects of the program environment have been ruled out or explained.

Applying Definition to Extension

This definition *implies* rather succinctly that program effects can be identified, measured, and separated from those of other origins. Cooperative Extension may experience difficulty in demonstrating all three conditions.

Identification

At least three problems are associated with the first condition, identification:

1. Many of Extension’s goals are so broad (to strengthen American families, to aid communities in developing and maintaining a satisfying

economic and social environment),⁹ that it's difficult to determine how related changes will manifest themselves (for example, what would be indicators of strengthening American families?).

2. Some programs are established with only process goals or no explicit goals at all.
3. Effects may be intended or unintended; positive, negative, or neutral in value; stable or unstable; immediate, intermediate, or long-term; and for primary, secondary, or tertiary participants.¹⁰

Identifying potential effects or outcomes to study isn't a simple matter and identification is confounded by determinations of what's of most import to study from the socio-political viewpoint.

Measurement

The second condition is that effects can be measured. A problem with social programs, and especially those that have been in operation for a long time, is that they tend to have effects that are small and difficult to measure.¹¹ Extension fits that description, plus it has the problem that the usual measures of educational experiences (tests, surveys) are inadequate for measuring what Extension hopes to accomplish with its educational offerings. Even if these effects could be easily quantified, many of the changes would need to be observed in environments outside the usual places and situations for data collection.

The questions being asked now about Extension programs, at least at state and national levels, are . . . generalizable questions. Funding groups want to make sure that reports on previously implemented programs will be repeated if they reinvest their scarce resources and that what was reported specifically is likely to occur generally.

Separation

The third condition is that Extension's effects could be separated from others. However, social phenomena are complex, they arise from many roots and causes and are influenced by circumstances other than those introduced by a program.¹² Ferreting out Extension program effects from the myriad of others is magnified by the number of other agencies working on the same or similar problems as Extension.

A recent example is energy conservation. Many other agencies have made conservation information available, plus other forces, economic and regulatory, influence deci-

sions to change. Similarly, individuals will seek information from a variety of sources before making decisions. They themselves may not know what sources were most significant in their decisions for change. Thus, sorting out Extension's contributions can be difficult and require a large output of resources.

Implications of the Definition

The previous discussion suggests that Extension may be unable to implement "impact evaluations" to meet the conditions of the definition—but that's not the case. The key is design—design of programs and design of evaluations.

Identifying Effects

The first condition of the definition was that effects can be identified; the associated Extension situation was goals stated in broad, vague terms, stated as process, or not stated at all. The remedy here is different depending on whether the program to be subjected to impact evaluation is in developmental stages or already ongoing.

If it's the former, the need is for a problem-focused program design. The first set of requirements is:

1. Awareness of a problem considered important to those with a vested interest in Extension (fundors, clientele, and other stakeholders).
2. Identification of the intensity and/or pervasiveness of the problem in a specific audience.
3. A hypothesis of what caused the problem to occur or what research suggests as a solution.
4. An understanding of when and why people change.

These requirements should lead to the identification of what needs to be accomplished and what's possible to do—clear and worthy goals and objectives.

Note that the focus here is on something specific that needs to be changed—a problem—rather than on some ultimate outcome or consequence. It's the difference, for example, between planning to increase profit from beef production (ultimate outcome) and planning to eradicate brucellosis (a problem), between planning to improve family health (ultimate outcome) and planning to decrease obesity (problem). This focus should make it easier to identify significant objectives and justify our programs to those outside Extension.

If the program is ongoing, impact evaluation shouldn't be tried unless goals/objectives are explicit enough to determine what would represent indications of accomplishment with an identifiable audience on a specific problem.

In some cases, unclear goals may be made explicit through discussions with program staff. A second alternative is to try to infer what the program is trying to achieve by observing it in operation. This clarification of program emphasis in hindsight can be time-consuming. But, it may suggest adjustments that could avert program failures that the usual evaluation of effects would require a great deal of time to detect and may avoid the waste of resources spent on evaluating a program that never got off the ground.

*Measuring
Effects*

The second condition of the definition was that effects can be measured; the problems in Extension were the probability of small effects and the inadequacy of measuring techniques. A partial remedy is again a problem-focused program design.

Before any aspect of a program is implemented, goals should be analyzed and a comprehensive list made of ways the audience would behave if the objectives had been met. Listing the indicators up front increases the probability of specific steps being taken to make the most important ones happen and enables measurements to be made at the most opportune times to show impact—not just at the end of the program.

The second requirement of a problem-focused design—identification of the intensity and/or pervasiveness of the problem in a specific audience—should help assist measurement in three ways: (1) provide benchmark data for comparisons of impact at later periods of time, (2) enable the design of measuring techniques that fit the audience's specific personal capabilities and lifestyles, and (3) decrease the underestimate of impact that's caused from the collection of data from inappropriate people.

Here's an example of the last point. Beef producers consist of at least three audiences:

1. Small landowners or hobbyists, most of whom work off the farm, who, with only a few head of cattle, can't afford to implement many of the production practices recommended to make a cattle operation a profitable business.
2. Large landowners holding land for appreciation, who have cattle only to qualify for the lower tax rate associated with agricultural assessment. Some may participate in an occasional Extension program, but they're not interested in spending money to improve their cattle operations.
3. Those who run cattle operations as economic enterprises. This group may be further categorized by operation type (specialized, diversified) and by operation size.

Extension should handle the requests of Groups 1 and 2, but proactive programs should probably be planned only for Group 3 and only this group be measured for impact. Too often, data are included from Groups 1 and 2 who can't or won't change (plus agribusinesses that aren't even in cattle production), which pulls down the average or overall impact measured for "beef producers." However, sociopolitical forces sometimes require that programs be planned for Group 1 even though the measurable economic impact may be negligible.

Increasing the planning system to four years should also improve Extension's ability to show impact. For problems and/or audiences that are slow to change, longitudinal studies—measurements made at year one and then at several four-year intervals—should pick up trends. And, slight change accumulated over time can sum to a considerable total, have valuable consequences for society,¹³ and be convincing evidence of program success.

*Distinguishing
Program
Effects*

The third condition of the definition was that program effects could be separated from effects from other sources. Extension's situation was that social problems with which we typically deal are complex in origin and influence and that other agencies make simultaneous efforts on these problems. The remedy is again problem-focused program design and evaluation design.

The purpose of evaluation design is to control or explain alternative reasons (other than the program) for effects to be found.¹⁴ Many people in Extension will argue that this purpose—establishing cause for effect—is too difficult or too demanding of resources for Extension to do. We agree that it takes extra effort, but argue that some control is quite do-able, plus many chances exist to explain if nonprogram events could have brought about measured change. For example, comparison groups could be set up when more people want to participate in a program than the staff can accommodate. The "extras" could serve as controls for the first group and then participate in the program at a later date.

Similarly, effects of other sources could be explained. For example, a serious competitor for effects of Extension programs is what's referred to as history or extraneous events.¹⁵ This means that program participants may be exposed to some other informational source on the same topic while the Extension program is ongoing.

This effect could be explained by program staff keeping close watch on what's occurring in the environment of the participants. For example, if the program focused on dietary habits of low-income young mothers, scrutiny should be

made of television programs, magazines (especially those sold at grocery stores), food stamp materials, newspaper articles, and any other source to which the group typically might have access.

When evaluations are conducted in multiple sites, the above-mentioned threats to valid findings of program effects are of less import. What's needed here is a design that will capture what naturally occurred at each site. It makes very little sense to hold programs constant over several sites for evaluation purposes when the natural condition in Extension is a great deal of variability in the same program over different sites.

In multiple natural sites, validity is established from accurate and detailed reports of the environments internal and external to the sites studied, which allows program staff in other settings to judge how close their own situations fit the ones where change was recorded.¹⁶

This discussion on evaluation design is important to impact studies, but less so when a problem-focused program design has been developed. For example, if a problem is identified in a specific audience and a program is developed that (1) has identifiable and measurable effects or outcomes, (2) has resource inputs and program activities that are appropriate to achieve the intended outcomes, and (3) has been delivered to the appropriate audience in sufficient degree and for a sufficient time for the change to be expected to occur, then it seems reasonable for Extension to claim major credit for the changes without having to factor out, absolutely, the other variables that may have contributed to the changes.

Other Implications

The decision to implement impact studies in Extension was in response to questions about program effectiveness being asked by people mostly outside Extension. For that reason, evaluating Extension programs for impact requires that data be collected and analyzed in systematic, purposeful ways. It requires that standardized procedures be used such that the findings of program results will be recognized and trusted by non-Extension types. It requires moving from "opinion" to "fact," from "feeling" that a program works to "having evidence" that it works. It makes little difference if data are qualitative or quantitative, of the so-called hard or soft variety, as long as the data collection is focused on specific questions, implemented uniformly with appropriately selected respondents, and analyzed with trusted procedures.

Too often, in Extension evaluation reports, only descriptive statistics of participants are used—how many, who, or what percentage did something. This isn't adequate for

impact studies. To determine effects of programs requires comparisons: programs against standards, against similar programs, against no program, or against the same program over time. Correlational and/or inferential statistics are then used to show relationships or differences between/among comparisons.

Conducting impact evaluations to provide defensible answers to the questions now being asked of Extension will require great resources. The best yield for these resources will come from evaluations designed for many sites and perhaps many states. A study showing small effects over a collective of sites is more credible evidence of program impact than any impact reported from a small, individual study.

To report that a farm management program in one county produced major economic consequences for one group of agricultural producers is good news, but it doesn't allow conclusions about farm management programs in other counties.

Extension programs are so varied from one county to another and so influenced by individual agent ability and initiative that the one-site studies are vulnerable to any number of competitive reasons for program results. However, if a number of different sites under natural, uncontrolled conditions, with different agents and different participants, produced effects all in the same direction, a plausible conclusion would be that the program contributed to those effects.

The questions being asked now about Extension programs, at least at state and national levels, are these generalizable questions. Funding groups want to make sure that reports on previously implemented programs will be repeated if they reinvest their scarce resources and that what was reported specifically is likely to occur generally.

Conclusion

Impact studies in Extension grew out of concerns for accountability to people mostly outside Extension and mostly for the purpose of continued funding of programs. To provide this accountability, that is, to answer others' questions, we inside Extension need to understand what the term "impact evaluation" means to outsiders, be aware of some of the problems we face in conducting impact studies, and start focusing our efforts where the greatest value will result.

Footnotes

1. P. H. Rossi and H. E. Freeman, *Evaluation: A Systematic Approach* (Beverly Hills, California: Sage Publications, 1982).
2. *State Extension Plan of Work and Report Guidelines: October 1, 1983-September 30, 1987* (Washington, D.C.: ES/USDA, 1983).

3. *Standards for Evaluation* [draft] (Washington, D.C.: Evaluation Research Society, 1980).
4. ERS Standards Committee, "Evaluation Research Society Standards for Program Evaluation," in *Standards for Evaluation Practice*, P. H. Rossi (San Francisco: Jossey-Bass, 1982).
5. M. Q. Patton, *Practical Evaluation* (Beverly Hills, California: Sage Publications, 1982).
6. Lee J. Cronbach, *Designing Evaluations of Educational and Social Programs* (San Francisco: Jossey-Bass, 1982).
7. Rossi and Freeman, *Evaluation: A Systematic Approach*.
8. J. R. Sanders, "Criteria for Technically Valid Impact Studies," in *Evaluation in Cooperative Extension Programs*, M. F. Smith, compiler (Proceedings from minisession at Evaluation Network Annual Conference, Baltimore, Maryland, 1982).
9. USDA/NASULGC Joint Committee, *Extension in the '80's: A Perspective for the Future of the Cooperative Extension Service* (Madison: University of Wisconsin-Extension, Cooperative Extension Service, 1983).
10. J. R. Sanders, "Measuring Consequences of Programs," in *Current Issues/Problems in Evaluating Cooperative Extension Programs*, M. F. Smith, ed. (Gainesville: University of Florida, Cooperative Extension Service, Institute of Food and Agricultural Sciences, 1981).
11. Rossi and Freeman, *Evaluation: A Systematic Approach*.
12. *Ibid.*
13. J. P. Gilbert, R. J. Light, and F. Mosteller, "Assessing Social Innovations: An Empirical Base for Policy," in *Evaluation and Experiment*, C. A. Bennett and A. A. Lunsdaine, eds. (New York: Academic Press, 1975).
14. Rossi and Freeman, *Evaluation: A Systematic Approach*.
15. D. T. Campbell and J. C. Stanley, *Experimental and Quasi-Experimental Designs for Research* (Chicago: Rand McNally, 1966).
16. Cronbach, *Designing Evaluations*.