

6-1-2009

## Theory and Rigor in Extension Program Evaluation Planning

Marc T. Braverman

*Oregon State University*, [marc.braverman@oregonstate.edu](mailto:marc.braverman@oregonstate.edu)

Molly Engle

*Oregon State University*, [molly.engle@oregonstate.edu](mailto:molly.engle@oregonstate.edu)



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

---

### Recommended Citation

Braverman, M. T., & Engle, M. (2009). Theory and Rigor in Extension Program Evaluation Planning. *The Journal of Extension*, 47(3), Article 1. <https://tigerprints.clemson.edu/joe/vol47/iss3/1>

This Feature Article is brought to you for free and open access by the Conferences at TigerPrints. It has been accepted for inclusion in The Journal of Extension by an authorized editor of TigerPrints. For more information, please contact [kokeefe@clemson.edu](mailto:kokeefe@clemson.edu).



**June 2009**  
**Volume 47 Number 3**  
**Article Number 3FEA1**

[Return to Current Issue](#)

# **Theory and Rigor in Extension Program Evaluation Planning**

**Marc T. Braverman**

Program Leader, Extension Family and Community Development Program  
Professor, Human Development and Family Sciences  
[marc.braverman@oregonstate.edu](mailto:marc.braverman@oregonstate.edu)

**Molly Engle**

Associate Professor and Extension Evaluation Specialist  
[molly.engle@oregonstate.edu](mailto:molly.engle@oregonstate.edu)

Oregon State University  
Corvallis, Oregon

---

**Abstract:** This article examines two aspects of evaluation planning for Extension programs: the use of program theory and logic models and the decision process that affects the evaluation's methodological rigor. First, Extension program planners should move beyond standard applications of logic modeling to incorporate a broader, more flexible use of program theory. Second, a highly rigorous evaluation will provide numerous benefits, but considering the costs that are typically required, an evaluation's degree of rigor should be carefully determined with reference to how the findings will be used. The article makes recommendations with the aim of promoting effective and economical Extension evaluations.

---

## **Introduction**

Embedded within a network of land-grant universities, county partnerships, and the USDA, Cooperative Extension is as complex an organizational system as you are likely to find. Due to this complexity, Extension has an unusually wide range of stakeholders-legislators, funders, clientele audiences, agency partners, academic colleagues, taxpayers, and others-who are interested in knowing the quality and effectiveness of our programs. Evaluation is the primary way that we provide that information, and, therefore, evaluation is critical to Extension's long-term organizational success.

Because of the range of evaluation activity across Extension, our evaluations should strive to be both effective and economical, answering the critical questions of stakeholders or program staff in a way that makes best use of available resources. This process begins in the planning stage. In this article we examine two connected elements of evaluation planning that can help us meet these goals: the program theory (and associated logic model) that forms the basis for deciding what the evaluation will cover and the rigor of the evaluation's methodology.

## Using Logic Models to Build More Sophisticated Program Theory

Our first general recommendation is for Extension program planners and evaluators to make more advanced use of program theory and logic models and thereby take advantage of the benefits of these practices for building effective programs. Both of these terms have been defined in multiple ways in the evaluation literature, but basically, a *program theory* is an explanation of how the program is supposed to work and how it will bring about the intended change for the program's target audience (Bickman, 1990; Rossi, Lipsey, & Freeman, 2004). A *logic model* is a visual representation—a diagram—of the basic elements of the program theory. Logic models can take different forms, but the distinctive format of listing columns labeled *Resources, Activities, Outputs, Outcomes, and Impacts* (or similar categories) has been popularized through efforts of numerous organizations and evaluators, both inside and outside of Extension (e.g., Bennett, 1975; United Way of America, 1996; University of Wisconsin-Extension, 2005; W.K. Kellogg Foundation, 2004).

Within Extension, logic modeling has achieved widespread adoption over the past 15 years, led by the pioneering efforts of Extension faculty at the University of Wisconsin, who have made extensive use of national trainings and online dissemination. As Taylor-Powell and Boyd (2008) note: "Today, the logic model forms the basis of the federal planning and reporting system and is widely used and adapted by Extension organizations for program planning, evaluation, reporting, and grantwriting purposes" (p. 65). Examples of logic models used as the basis for programmatic research can also be found in the *Journal of Extension* (e.g., Hosty, 2005; Schmidt, Kolodinsky, Flint, & Whitney, 2006).

As many Extension staff know, logic models can help us determine where to focus an evaluation. But they have other important uses as well (Morell, 2008). They can help us to analyze program shortcomings and guide program improvements; they are excellent educational tools for describing our programs to stakeholders; and they can strengthen our grant proposals.

### Moving Beyond the Standard Logic Model

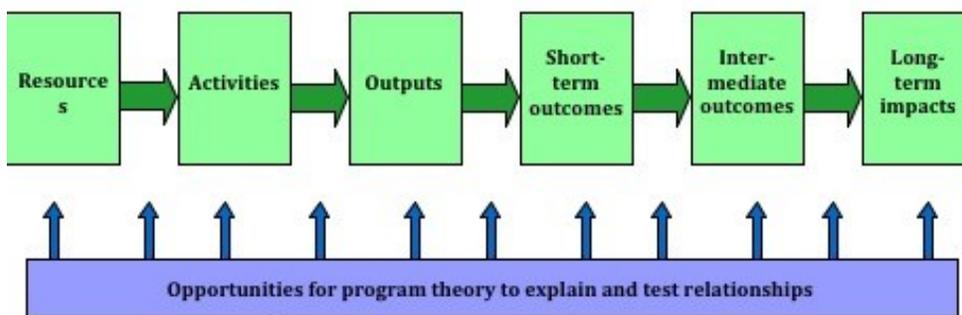
With this background of success, we think the time is right for Extension professionals to move beyond the standard logic model framework into new territory. Much of the potential of the logic modeling concept has not yet been tapped in Extension, and we can incorporate program theory in more sophisticated ways. Here are several examples of how this might occur:

- *More specific causal connections.* The standard format calls for a long list of elements within each box. The list of program activities leads to a list of outputs, which leads in turn to a list of short-term outcomes, longer-term impacts, etc. The causal arrows flow from one box to the next, and it is not possible to tell which elements within the boxes are necessary for change to take place. A more incisive, informative model would specify the individual causal connections between these elements. For example, which specific activities does a particular learning outcome depend on? This would create a more complex logic model, but one that is also more useful for understanding and testing the program.
- *Focusing evaluation activity on selected parts of the model.* The program theory underlying a logic model may be relatively better established in some sections of the model than in others. In those cases, the program theory can guide decisions about which parts of the model raise the biggest concerns and deserve the most evaluative attention. For example, Dale Blyth (2009) has recommended an overarching evaluation model for 4-H youth development programs in which at the

program level, Extension personnel focus on assessing the *quality* of program implementation rather than outcomes, while the linkages between program quality and the intended outcomes are measured at a more centralized level of Extension. Clearly, this approach to evaluation requires a careful reliance on program theory to make the different pieces fit together.

- *Identifying critical mediators within any section of the logic model.* Part of the value of program theories and logic models is that their causal links can help identify potentially important mediators—those intermediate variables that are necessary to make the desired change happen. In most standard logic models, mediating relationships are included only in the chain of outcomes: the program creates short-term *knowledge* change, which leads to intermediate-term *behavior* change, which finally leads to long-term *impact* such as improved personal health or economic sustainability. In other words, the causal relationship between the program operation and the long-term impact is mediated by changes in both knowledge and behavior. However, as we will illustrate shortly, mediating relationships can occur at any point in the logic model, not just among the outcomes. A solid program theory can illuminate details of the program process at any point in the logic model. Figure 1 illustrates this concept diagrammatically.

**Figure 1.**  
Using Program Theory to Expand the Basic Logic Model



## A Real-Life Example: The Oregon 4-H Horse Program

An illustration will show how these concepts can be applied in the evaluation planning process. Figure 2 shows the Oregon 4-H Program model, displayed as a logic model, somewhat modified from the standard format. Working backward, the model shows that positive youth development—as exemplified by the variables of competence, confidence, connection, etc.—will occur if the program can successfully provide opportunities for youth to experience belonging, mastery, generosity, and independence. Those opportunities, in turn, will be made possible by a combination of adequate resources, youth participation, and engagement strategies such as project meetings.

State 4-H faculty have recently begun studying the 4-H horse program as a case study, to see how these principles work in action, and they used program theory to delve into a deeper understanding of program operations. Staff members had heard anecdotally that the horse project was frequently characterized by high conflict among adults, both volunteer leaders and parents, within its local program units. The staff feared that this conflict could potentially interfere with the successful achievement of program outcomes, as detailed in the logic model. Therefore they embarked on a small, qualitative study of the role of conflict, with the goal of

building a stronger program theory that could, in turn, guide a comprehensive program evaluation.

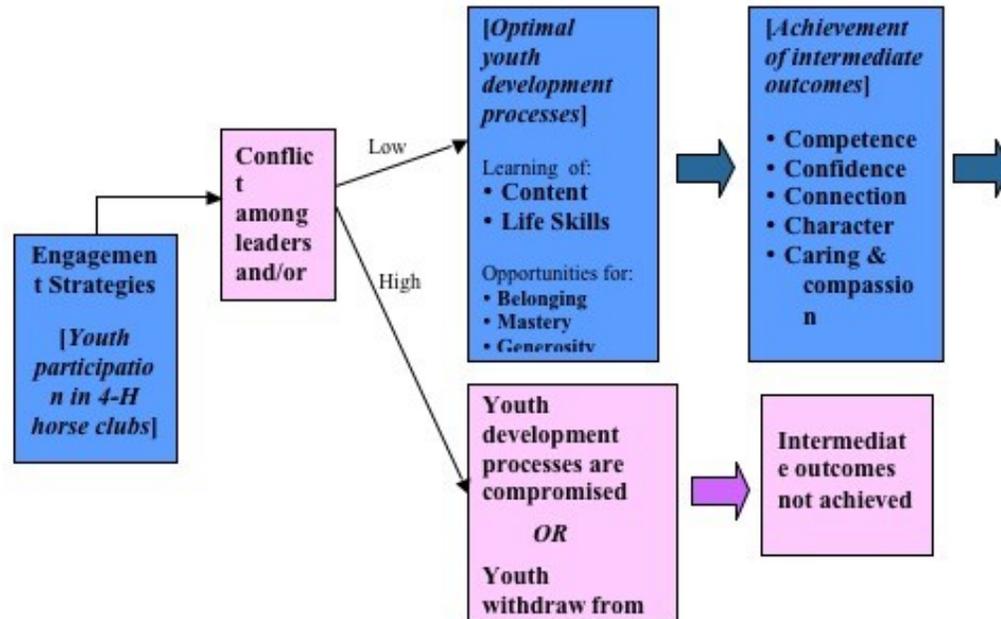
**Figure 2.**  
The Oregon 4-H Program's Logic Model for Educational Activities



The 4-H staff conducted a series of focus groups around Oregon, which suggested that conflict did indeed seem to exist in several, though certainly not all, 4-H horse clubs. They wondered whether the level of conflict within a club might be an impediment to the program's ability to achieve successful youth outcomes, and they began to consider this as a potential new element for the horse program's program theory. The presence of conflict between adults might mediate the program's ability to move successfully from engagement strategies to learning opportunities. These new *potential* relationships are illustrated in Figure 3. The model is oversimplified since, of course, the conflict variable is more complex than the simple dichotomy *low-high*. An evaluation will need to take account of this greater real-world complexity, relying on aspects of the program theory that are not represented in the simplified logic model diagram (that is, how to conceive and measure *conflict*, a challenge that is also relevant for our next section).

Currently, the 4-H program staff are in the process of developing an evaluation of these very relationships. If they do find that conflict is an impediment for achieving the intended positive youth development outcomes, the staff will be in a position to address the problem directly and search for solutions. For our illustrative purposes, we can see that the program theory and logic model have been expanded in a way that allows for more sensitive examination of the program's interpersonal dynamics. This may set the stage for refining the horse program to keep it on track, and to avert a potentially serious roadblock to success for its youthful clientele.

**Figure 3.**  
Section of Logic Model for a Program Theory Hypothesis About How Conflict Within the 4-H Horse Program Might Interfere with Educational Processes



## Methodological Rigor: Building a Case Based on Persuasive Evaluation Evidence

We have all heard the call. Funding agencies, grant reviewers, legislators, and our academic departments desire methodologically rigorous evaluations of Extension programs, that is, evaluations that are technically sound and provide an opportunity to show solid, convincing evidence of a program's impact. This point is made within the "pages" of *JOE* as well (e.g., Duniform, Duttweiler, Pillemer, Tobias, & Trochim, 2004). The rigor, or technical strength, of an evaluation can influence whether a program is continued or eliminated, or whether it achieves recognition and distinction.

### Methodological Rigor Defined

Braverman and Arnold (2008) define rigor as "a characteristic of evaluation studies that refers to the strength of the design's underlying logic and the confidence with which conclusions can be drawn" (p. 72). A rigorous evaluation makes use of strong designs and valid measures and tracks information about how the program was actually delivered. Rigor contributes to evaluation quality, and it can be described in terms of specific elements related to the evaluation's planning and implementation. Several of those critical elements are the following (see Braverman and Arnold for a more detailed list):

- *Evaluation design*: For program impact evaluations, how well does the design allow us to determine if the program itself was the cause of positive change in the outcomes?
- *Measurement strategies*: Will the program outcomes be measured in a valid, reliable way that provides strong evidence for drawing conclusions?
- *Program monitoring*: During the evaluation, are we observing the program closely enough so that we can describe how it is being delivered, including potential differences between program delivery sites?

- *Program participation and attrition:* Are efforts made to reach participants who didn't attend regularly, who left the program midway, or who received different program dosage levels? Or does it just measure whoever happens to attend on the day of data collection?

For most of these elements, a number of options will probably exist for the evaluation planning team. Those options may range from high rigor-producing high confidence in the findings-to a more moderate level of rigor, which may be somewhat less convincing but will still allow for valid and useful conclusions. We don't include a "low-rigor" category because that would threaten the basic acceptability of the evidence and lead to a weak—and not particularly useful—evaluation.

Table 1 displays these concepts in action. Using the example of an educational program aimed at teaching parenting skills, it shows some of the many choices that can be made during the evaluation-planning period. In each case, a moderate-rigor choice is contrasted with a high-rigor choice, and the advantage of the high-rigor choice is described.

**Table 1.**

Options in Evaluation Planning: Moderate-Rigor and High-Rigor Choices to Evaluate a Multiple-Session, Multiple-Site Parenting Skills Education Program

<b>Rigor Element</b>	<b>Moderate-Rigor Option</b>	<b>Higher-Rigor Option</b>	<b>What the Higher-Rigor Option Adds</b>
Evaluation study design	Single group pre- and post-test design	Comparison group design	More confidence that our program was the cause of positive change (if indeed positive change occurs)
Measurement - Knowledge gain	Participants' self-ratings of how much they learned about good parenting	Valid, reliable test of what people actually know about the program's content	Being able to make more authoritative statement about what people really know (and don't know) after participating in the program
Measurement - Behavioral change	Participants' intentions (at end of class) to change their parenting behaviors	Six months after program, self-report surveys of participants' current parenting behaviors	More confidence in stating that the program has resulted in actual behavioral change
Program delivery monitoring	Observe one session per delivery site, or interview program leader to determine what content was	Observe multiple sessions at each delivery site, to get a detailed picture of program delivery	Ability to explain, rather than speculate, why delivery sites may differ from each other in effectiveness

	covered		
Program participation and attrition	Give survey to only those participants who attend the final class session	Program team determines beforehand what minimum number of sessions should count for program participation, and makes attempt to survey an appropriate sample of participants who meet that attendance standard	More comprehensive understanding of the program's full audience, rather than a convenience sample of people who attended on a given date

## The Benefits and Costs of Methodological Rigor

What are the benefits of a stronger evidence base and greater confidence in your evaluation results? Here are a few.

- A strong evaluation can make a program more competitive for obtaining follow-up funding from government agencies or private foundations.
- It can help a program gain recognition at the national level.
- It can help convince elected officials of Extension's value (especially in combination with the two bullets above).
- It will make it possible to disseminate the program and the evaluation study via refereed publications, conferences, and other high-visibility channels.
- Through the program's increased recognition, a strong evaluation can help attract new participants and help the program grow.
- It can lead to increased understanding among Extension staff of the program's true strengths and shortcomings, and a clearer sense of what needs to be done to strengthen the program.

Keeping these benefits in mind, however, there is another side to the coin. Program evaluation can be an expensive undertaking to begin with, and the factors associated with strong methodology can increase costs quickly. For example, consider these sources of increased expense for an impact evaluation:

- Recruiting a large evaluation sample to increase the study's statistical power,
- Recruiting, testing, and monitoring a comparison group that does not receive the program,

- Extensive monitoring of program sessions,
- Using multiple measurement points (e.g., three rather than two, or two rather than one),
- Compensating evaluation study participants in order to minimize attrition,
- Finding professionals with the expertise to develop instruments or conduct specialized statistical analysis, and
- Tracking long-term behavioral outcomes in addition to short-term learning outcomes.

All of these factors require time, money, and personnel. The sum total may exceed your program's budget, time horizon, or expertise. Therefore, although a moderate-rigor evaluation will result in only a moderate level of confidence in the findings, it will be easier to implement, will require fewer resources, and may take less time.

## Making Choices

As can be seen, the rigor of an evaluation is a multidimensional concept rather than a unitary one. It's an accumulation of different parts, each requiring care and attention. We have seen cases of impact studies that have strong comparative designs, such as separate program and control groups, but weak or inappropriate measurement strategies. Unfortunately the net effect is an evaluation that is not very convincing, despite the impressive control group design. Evaluation teams have to make smart, thoughtful, and sometimes difficult decisions about methodology, weighing benefits and costs to find the most suitable overall plan.

How should program staff and the evaluation team make these choices? It is frequently said that you should plan the most rigorous evaluation possible for the amount of resources available. To that commonplace bromide, we add two caveats. First, if you are indeed working with a fixed amount of resources-by which we are generally referring to money and people's time-there are still plenty of decisions about whether to put those resources into one part of the evaluation or another.

- Which is a more critical use of staff time: observing numerous program sessions or conducting additional focus groups?
- Which is a more critical use of limited funds: purchasing well-tested, copyrighted instruments or compensating respondents to increase participation rates in the data collection?

These questions must be answered with the big picture in mind-how the evaluation information will be interpreted and used.

Second, these decisions require consultation with a wider group than just the program staff and evaluators. To find out if the evaluation evidence will be convincing, make sure you know the opinions of the people who need to be convinced-the evaluation's primary intended users (Patton, 2008). Decisions about whether a planned evaluation is sufficiently rigorous and whether the strength of evidence will be "good enough," need input from stakeholders, administrators, academic colleagues, and other individuals who will be making

critical decisions. Some evaluation situations, e.g., building a program from a federal grant, will require the highest methodological rigor possible, despite the costs. In other situations, a more moderate level of confidence may be acceptable, and it may be important to limit the resources expended. In those cases a moderate-rigor evaluation will often be an appropriate and preferred option.

## Conclusion

In this article we have offered some recommendations in two specific areas—using program theory and deciding on methodological rigor—to help Extension professionals plan evaluations that are both effective in targeting information needs and economical in the use of resources. Some final summary points to keep in mind are the following.

- Extension program planners should move beyond the standard applications of logic models—helpful as they are—into a broader, more flexible use of program theory.
- Attention to program theory can help to identify areas that need examination and to focus on high-priority evaluation questions.
- The rigor of a program evaluation depends on a combination of specific planning decisions and implementation factors.
- The value of highly rigorous evaluations lies in the confidence one can place in the findings and conclusions: the more rigorous the evaluation, the more likely it will be able to stand up to critical scrutiny. However, rigor can also involve higher costs in terms of time and money.
- The many planning decisions relating to rigor should be made with awareness of the plan for using the findings, as well as attention to available resources. A moderately rigorous evaluation can be appropriate if it will answer the primary questions with a level of confidence that is acceptable for the program's stakeholders.

### Acknowledgments

We thank Mary Arnold, Roger Rennekamp, and Jonathan Morell for their valuable contributions to the development of this article.

## References

Bennett, C. (1975). Up the hierarchy. *Journal of Extension* [On-line], 13(2). Available at: <http://www.joe.org/joe/1975march/1975-2-a1.pdf>

Bickman, L. (Ed.). (1990). *Advances in program theory. New Directions for Program Evaluation*, 47.

Blyth, D. (2009, January). *Constructing the Future of Youth Development: Four Trends and the Challenges and Opportunities they Provide*. Webinar broadcast, 4-H National Learning Priorities conference. Retrieved April 20, 2009 from: <http://www.uvm.edu/extension/youthdevelopment/?Page=presentation-january09.html>

Braverman, M. T., & Arnold, M. E. (2008). An evaluator's balancing act: Making decisions about methodological rigor. In M.T. Braverman, M. Engle, M.E. Arnold, & R.A. Rennekamp. (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation, 120*, 71-86.

Dunifon, R., Duttweiler, M., Pillemer, K., Tobias, D., & Trochim, W. M. K. (2004). Evidence-based Extension. *Journal of Extension* [On-line], 42(2) Article 2FEA2. Available at: <http://www.joe.org/joe/2004april/a2.php>

Hosty, M. (2005). 4-H Wildlife Stewards-A new delivery model for 4-H. *Journal of Extension* [On-line], 43(5) Article 5IAW3. Available at: <http://www.joe.org/joe/2005october/iw3.php>

Morell, J. A. (2008, November). *Logic models: Uses, limitations, links to methodology and data*. Workshop presented at the annual meeting of the American Evaluation Association, Denver.

Patton, M. Q. (2008). *Utilization-focused evaluation (4th ed.)*. Thousand Oaks, CA: Sage.

Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2004). *Evaluation: A systematic approach* (7th ed.). Thousand Oaks: Sage.

Schmidt, M. C., Kolodinsky, J. M., Flint, C., & Whitney, B. (2006). The impact of microenterprise development training on low-income clients. *Journal of Extension* [On-line], 44(2) Article 2FEA1. Available at: <http://www.joe.org/joe/2006april/a1.php>

Taylor-Powell, E., & Boyd, H. H. (2008). Evaluation capacity building in complex organizations. In M.T. Braverman, M. Engle, M.E. Arnold, & R.A. Rennekamp (Eds.), *Program evaluation in a complex organizational system: Lessons from Cooperative Extension. New Directions for Evaluation, 120*, 55-69.

United Way of America. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: Author.

University of Wisconsin-Extension. (2005). *Logic model*. Retrieved April 20, 2009 from: <http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html>

W.K. Kellogg Foundation. (2004). *Logic model development guide*. Retrieved April 20, 2009 from: <http://www.wkkf.org/Pubs/Tools/Evaluation/Pub3669.pdf>

---

*Copyright* © by *Extension Journal, Inc.* ISSN 1077-5315. Articles appearing in the Journal become the property of the Journal. Single copies of articles may be reproduced in electronic or print form for use in educational or training activities. Inclusion of articles in other publications, electronic sources, or systematic large-scale distribution may be done only with prior electronic or written permission of the *Journal Editorial Office*, [joe-ed@joe.org](mailto:joe-ed@joe.org).

If you have difficulties viewing or printing this page, please contact [JOE Technical Support](#).