Program Theory and Quality Matter: Changing the Course of Extension Program Evaluation

Abstract

As internal evaluators for the 4-H program in two states, we simultaneously yet independently began to change the way we approached our evaluation practices, turning from evaluation capacity building (ECB) efforts that prepared educators to define and measure program outcomes to strategies that engage educators in defining and measuring program quality. In this article, we discuss our similar experiences, how these experiences are changing the ECB work we do, and how changing our evaluation approach ultimately will position 4-H for better evaluations in the future. This shift in evaluation focus has implications for other Extension program areas as well.

Mary E. Arnold
Professor and Youth
Development
Specialist
Oregon State
University
Corvallis, Oregon
mary.arnold@oregonst
ate.edu

Melissa Cater
Assistant Professor
and Program
Evaluation Specialist
Louisiana State
University
Baton Rouge,
Louisiana
mcater@agcenter.lsu.
edu

Introduction

Extensive research conducted by the Forum for Youth Investment (Smith, et al., 2012) demonstrated that evaluating program quality, as opposed to outcomes, leads to continuous program improvement and, ultimately, to more effective programs. However, focusing evaluation on measuring program quality rather than outcomes is a radical departure from the outcomes measurement approach that is currently prominent in Extension (Lamm, Israel, & Diehl, 2013).

Program evaluation—how to do it and how to help Extension get better at it—is a persistent topic in Extension-related literature. In their reflection on the state of the past, present, and future of the 4-H youth development program, Borden, Perkins, and Hawkey (2014) note that accountability is severely lacking and that 4-H needs to lead the way in using innovative evaluation methods to determine program effectiveness and improve program quality. The authors state that it is not just program outcomes that need measuring but also program quality and that the lack of standardized program implementation must be addressed before any major evaluation effort can occur.

The call for action by Borden et al. (2014) reflects the quiet but pervasive shift that is taking place in

youth development program evaluation. This shift is a movement away from measuring program outcomes without first establishing a clear understanding of what is happening at the program development and implementation level, and it has implications for the evaluation capacity building (ECB) efforts that are required. While Ghimire and Martin (2013) found that the 4-H program has greater levels of ECB efforts already in place than other Extension program areas, the emphasis of these efforts remains on skills related to outcomes evaluation.

As internal evaluators for the 4-H program in two very different states, we simultaneously yet independently began to change the way we approached our evaluation practices, turning from ECB efforts that prepared educators to define and measure program outcomes to strategies that engage educators in defining and measuring program quality. In this article, we discuss our similar experiences, how these experiences are changing the ECB work we do, and how changing our evaluation approach ultimately will position 4-H for better evaluations in the future. It is important to note at the outset that we are not proposing abandoning the measuring of program outcomes. Rather, we propose shifting ECB efforts toward (a) the articulation of sound program theory, which by its very nature includes defining program outcomes; (b) attending to program quality and what that means for program implementation; and (c) moving outcomes measurement to a higher level in the organization. This shift in evaluation focus has implications for other Extension program areas as well (Arnold, 2015). To set the stage for recommending broad implementation of revised evaluation practices, we present a brief history of ECB efforts in Extension and provide a review of the program quality and theory of change literature that is the main driver for the change in our ECB approach.

ECB in Extension

Bennett (1975) is often credited with providing the first framework for conceptualizing Extension program evaluation. Referred to colloquially as "Bennett's hierarchy," this framework provided a clear and accessible description of the many levels of program evaluation in Extension, beginning with simple measures related to the resources invested in a program, moving to measurement of program participation and participant satisfaction with the program, progressing further to measurement of changes in participant learning and action, and ending with measurement of changes at the societal level. By the mid-1990s, Bennett's model was the cornerstone of nascent Extension ECB efforts and was included in Extension education training materials (Seevers, Graham, Gamon, & Conklin, 1997). Although Bennett's framework provided a way to describe and organize ECB needs and directions, it lacked any emphasis on articulating, let alone measuring, program theory and quality, which are the key elements that link program plans to outcomes.

Federal drivers in the form of the 1993 Government Performance and Results Act (GPRA), which focused attention on accountability for publically funded programs, and the 1998 Agriculture, Research, Extension and Education Reform Act, which extended GPRA to require plans of work and annual reports that demonstrated the achievement of medium-term outcomes and long-term impacts from Extension programs, set the stage for a singular focus on outcome measurement as the only acceptable evidence of Extension program impact. Although many Extension evaluators celebrated the movement away from measuring program participation and satisfaction as evidence of program success, the jump to measuring program outcomes left discussions of program quality

and theory, and their critical relationship to program outcomes, out of ECB efforts.

By the early 2000s, Bennett's framework was reflected in program logic models emerging in other organizations (Knowlton & Phillips, 2009; W. K. Kellogg Foundation, 2004). Extension easily made the switch from Bennett's framework to logic models, primarily due to the extensive ECB efforts led by the team at University of Wisconsin (UW) Cooperative Extension. Extension services across the country invested heavily in training, providing many educators the opportunity to attend one of the UW logic model training sessions. Within a few years, "inputs, outputs, and outcomes" became common Extension nomenclature, and state and federal program planning and reporting systems were developed based on logic modeling. Throughout the 2000s, Extension evaluators across the country focused ECB efforts on logic model training, preparing educators to think through the "logical" connections between what they did in their programs and the outcomes they hoped to achieve.

What followed the linear logic model approach was an emphasis on outcomes, particularly on how what participants learned in programs (short-term outcomes) translated into new actions (medium-term outcomes) with the hope that long-term societal changes would follow. This approach emphasized a logical connection between the program and its intended outcomes that *implies* program theory and a predictive intention (McLaughlin & Jordan, 2004). However, as Chen (2004) states, program success depends on the accuracy of the program's assumptions regarding these logical connections, and, therefore, the success of the program is dependent on the validity of those assumptions.

A recent study on current evaluation practices across the Extension system revealed that although many states use logic models for program planning and evaluation, evaluation efforts have been only minimally enhanced (Lamm et al., 2013). Furthermore, the majority of Extension evaluations involve a postprogram-only evaluation design. Nonetheless, the authors advocate for the continuation of ECB efforts based on logic modeling, without mentioning attention to program quality or theory, thus perpetuating the pervasive, if understated, Extension assumption that logic models automatically lead to sound outcomes. Implicit in this line of thinking is that the measurement of the stated outcomes automatically provides valid evidence of a program's impact, whether or not there is any theoretical connection between the program's activities and the outcomes and regardless of the quality of the program's implementation.

Theory of Change and Program Quality

One dilemma that faces the Extension system is building widespread understanding of the granularity of logic model versus theory of change. Patton (2002) differentiates logic models and theories of change by pointing out that the purpose of a logic model is simply to *describe*, whereas a theory of change is both *explanatory* and *predictive* because causal links in the program can be both hypothesized and tested. In addition to testable causal links, strong theories of change exhibit attributes such as social and organizational meaningfulness, plausibility, and obtainability (Hunter, 2006).

In simpler terms, a theory of change is defined as a program planner's "knowledge and intuition of what works" (Monroe et al., 2005, p. 61) in moving program participants to change. In many cases,

well-established theories, such as social learning, stages of change, and empowerment theory, are useful for Extension work (University of Maryland Extension, 2013). Human change may occur at a basic level of knowledge or skill, a more complicated level of attitudes or motivations, or a complex level of behavior change. A program's theory of change model explicates how these outcomes occur and may also identify mediators that are necessary for change to occur (Braverman & Engle, 2009). Theories of change offer value in various ways:

> Theories of change are valuable to program planners and stakeholders because they contribute to a common definition and vision of long-term program goals, how those goals can best be reached, and what critical program qualities need to be monitored and evaluated for successful achievement of the long term outcomes. (Arnold, Davis, and Corliss, 2014, p. 97)

Hunter's (2006) concept of theory of change incorporates both the micro, or program, level and the macro, or organizational, level with a strong focus on organization growth and financial sustainability. His view of the theory of change triumvirate incorporates an important program-level emphasis: program quality. This emphasis on program quality aligns with Blythe's (2009) suggestion that local-level programs should focus on improving the quality of the program, both the implementation of the program and the quality of the staff delivering the program: "Accountability for quality not outcomes is the key lever for improving impact at the program level" (Blythe, 2009, slide 14). Blythe also suggests that the responsibility for, and assessment of, outcomes rests at a geographic or policy level.

What does program quality mean, though? It is often a value-laden construct (Benson, Hinn, & Lloyd, 2001). The parties involved in defining quality, from stakeholders to evaluators, bring different sets of values to bear on the definition. In the absence of research to guide the process, consensus building is an important phase of defining program quality. Youth development programs that include Extension 4-H, however, have benefited from the recent focus on defining and measuring program quality. Program characteristics such as setting, engagement (i.e., breadth, intensity, depth, frequency, and duration), and supportive relationships between youth and adults (Eccles & Gootman, 2002) have been identified as highly important to program quality. Furthermore, program quality can be measured with valid and reliable instruments such as the Youth Program Quality Assessment (YPQA) tool (Smith & Hohmann, 2005). One strength of the YPQA is the identification of indicators of quality for each characteristic, using a rubric to indicate the degree of presence or absence of an indicator. As quality indicators are selected, the value piece becomes evident through the consensus-building process that involves both local educators and evaluators (Bickman & Peterson, 1990).

In the absence of research about a program's theory, some common points of entry can be used in describing program quality. Specifically, these points focus on the relationship of program structures and processes with participant outcomes (Baldwin & Wilder, 2014). Program structures include support features such as funding levels, staffing structure, and physical environment, whereas program processes involve delivery attributes such as educational activities, relationships among

educators and participants, and degree of participant engagement.

Within the Extension system, program quality is often best understood as commonly agreed on key program characteristics. A perusal of existing literature suggests that program quality is an issue that has not been clearly addressed across the Extension system. So often the discussion of program quality is general and may not even clearly connect to understanding how the quality of structures or processes impacts outcomes. On those occasions when it is discussed, the subsequent step of developing indicators of quality is rarely addressed.

Changing the ECB Approach—One Evaluator at a Time

As aforementioned, we are internal evaluators with the 4-H program in our respective states, and we recently discovered that we had both been changing our approaches to ECB. The changes we are making are strikingly similar and driven by the same three primary forces.

First, as our earnest efforts to build evaluation capacity unfolded, we both sensed that there was a critical tipping point. What is the appropriate evaluation capacity of Extension educators? And how does this expertise fit with all the other aspects of the increasingly demanding job of an Extension educator? We have witnessed issues of retention as educators struggle with expectations to be skilled at program management, principles of youth development, and program development—including program evaluation. We have watched the toll as increasing levels of stress and an inability to cope with the demands of the job cause employees to leave within the first three years. This led us to question whether our current approach to ECB was the most useful for the overall benefit of the 4-H program.

Second, changes in the structure of Extension appointments in some states have created a movement from tenure-track opportunities with a scholarship requirement to fixed-term educator appointments. Without the scholarship driver, there is little motivation for educators to develop evaluation skills beyond what is needed for reporting. As needed, we provide field educators some support with their local evaluations. However, most of our outcomes evaluation efforts now involve designing evaluations at the statewide level. Such evaluations are connected to statewide strategic goals and outcomes, which in the end are more useful for describing overall program impact. Local educators can participate by contributing data. In many cases, educators who contribute data to a statewide effort also receive a report based just on their own programs. Instead of traveling to build capacity, we spend more time analyzing statewide data and writing local and statewide reports. In doing so, we have lifted a great deal of the burden of program evaluation off the backs of overwhelmed educators.

Third, the utility of the evaluation results that were generated from our original ECB efforts was questionable because they were typically locally based evaluations of small programs that were difficult to aggregate with other similar evaluations. The results were rarely robust or conclusive enough to be generalized, and the postprogram-only design to which most educators defaulted, despite our efforts to broaden the design tool kit, were too simple to provide sufficient evidence of program impact, leaving us discouraged by how little sound impact evidence we had compared to the ECB effort that was made.

As we stumbled along in relative isolation, we began to rethink 4-H evaluation, moving from building capacity for outcomes evaluations with limited utility to the evaluation of organizational effectiveness based on sound program theory. As such, our ECB efforts are now targeted more toward actively engaging educators in evaluation at a level that they understand best: quality control of program structures and processes. And in doing so, we help educators understand and articulate the program theory so that they better understand the components of program quality and what they need to be doing at the local level for the program to be successful.

Implications of Changing ECB Focus

Changing the focus of ECB efforts has resulted in the centralization of the outcomes evaluation workload (Blythe, 2009), which reduces at least one expectation of field-based educators. In the process, however, we both have experienced less buy-in from educators and less interest in evaluation overall. Although evaluation opportunities are statewide, participation in these opportunities has not been as robust as it should be given the relative ease of participation for educators and the direct benefit they receive for participating. Over time, however, we are witnessing a steady increase in participation, as educators understand the process better and experience the benefits of participating in statewide outcomes evaluation.

Devoting less of our time to ECB efforts related to developing logic models for local programs has meant that we have more time to think about program theory and program quality and to build ECB efforts around program quality. In Oregon, for example, the 4-H program is working with the Weikert Center for Youth Program Quality to train educators on program quality measurement. This innovative effort requires time and a change in ECB focus. However, as a result, we are making strides in articulating program theory for 4-H and designing an accompanying evaluation plan that will focus not just on program outcomes but also on the steps along the way that make the outcomes possible.

In our quest to demonstrate outcomes we have drifted far from sound program development. Educators often are so caught up in delivery program activities that they seem no longer to understand why they do the things they do or how to design a set of linked activities that contribute to a common outcome. We believe it is time to build consensus around what quality structures and processes in 4-H look like. To us, this focus is more fruitful than the hours we spent on past ECB efforts.

The focus on program theory and quality has the potential to engage educators in program evaluation at a level they understand better, and at a level where they have direct influence: the quality control of program structures and processes. Doing so will set the stage to respond to the call to action for meaningful accountability set forth by Borden et al. (2014), and we believe the same call to action has implications across all Extension programs.

References

Arnold, M. E. (2015). Connecting the dots: Improving Extension program planning with program umbrella models. *Journal of Human Sciences and Extension, 3*(2), 48–67.

Arnold, M. E., Davis, J., & Corliss, A. (2014). From 4-H international youth exchange to global citizen: Common pathways of ten past program participants. *Journal of Youth Development, 9*(2), Article 140902RS001.

Baldwin, C., & Wilder, Q. (2014). Inside quality: Examination of quality improvement processes in afterschool youth programs. *Child and Youth Services, 35*, 152–168. doi: 10.1080/0145935X.2014.924346

Bennett, C. F. (1975). Up the hierarchy. *Journal of Extension* [Online], *13*(1). Available at: http://www.joe.org/joe/1975march/1975-2-a1.pdf

Benson, A., Hinn, D. M., & Lloyd, C. (2001). Preface. In D. Michelle Hinn, Claire Lloyd, & Alexis P. Benson (Eds.) *Advances in program evaluation: Vision of quality: How evaluators define, understand, and represent program quality* (ix-xii). United Kingdom: Emerald Group Publishing Limited.

Bickman, L., & Peterson, K. (1990). Using program theory to describe and measure program quality. *New Directions for Program Evaluation, 47*, 61–72. doi: 10.1002/ev.1555

Blythe, D. (2009). Constructing the future of youth development: Four trends and the challenges and opportunities they provide [PowerPoint slides]. Retrieved from https://www.uvm.edu/extension/youthdevelopment/daleblythslides.ppt

Borden, L., Perkins, D. F., & Hawkey, K. (2014). 4-H youth development: The past, the present, and the future. *Journal of Extension* [Online], *52*(4) Article 4COM1. Available at: http://www.joe.org/joe/2014august/comm1.php

Braverman, M., & Engle, M. (2009). Theory and rigor in Extension program evaluation planning. *Journal of Extension* [Online], *47*(3) Article 3FEA1. Available at: http://www.joe.org/joe/2009june/a1.php

Chen, H. (2004). *Practical program evaluation: Assessing and improving planning, implementation, and effectiveness.* Thousand Oaks, CA: Sage.

Eccles, J., & Gootman, J. (Eds.). (2002). *Community programs to promote youth development.* Washington, DC: National Academy Press.

Ghimire, N., & Martin, R. (2013). Does evaluation competence of Extension educators differ by their program area of responsibility? *Journal of Extension* [Online], *51*(6) Article 6RIB1. Available at: http://www.joe.org/joe/2013december/rb1.php

Hunter, D. (2006). Using a theory of change approach to build organizational strength, capacity and sustainability with not-for-profit organizations in the human services sector. *Evaluation and Program Planning, 29*, 193–200. doi: 10.1016/j.evalprogplan.2005.10.003

Knowlton, L. W., & Phillips, C. C. (2009). *The logic model guidebook: Better strategies for great results.* Thousand Oaks, CA: Sage.

Lamm, A. J., Israel, G. D., & Diehl, D. (2013). A national perspective on the current evaluation

activities in Extension. *Journal of Extension*, [Online], *51*(1) Article 1FEA1. Available at:
http://www.joe.org/joe/2013february/a1.php

McLaughlin, J. A., & Jordan, G. B. (2004). Using logic models. In J. S. Wholey, H. P. Hatry, & K. E.
Newcomer (Eds.), *Handbook of practical program evaluation* (2nd ed.) (pp.7–32). San Francisco,
CA: John Wiley and Sons.

Monroe, M., Fleming, M., Bowman, R., Zimmer, J., Marcinkowski, T., Washburn, J., & Mitchell, N.
(2005). Evaluators as educators: Articulating program theory and building evaluation capacity. *New
Directions for Program Evaluation, 108,* 57–71. doi: 10.1002/ev.171

Patton, M. Q. (2002). *Qualitative research and evaluation methods.* Thousand Oaks, CA: Sage.

Seevers, B., Graham, D., Gamon, J., & Concklin, N. (1997). *Education through Cooperative
Extension.* Albany, NY: Delmar Publishers.

Smith, C., Akiva, T., Sugar, S., Lo, Y. J., Frank, K. A., Peck, S. C., Cortina, K. S., & Devaney, T.
(2012). *Continuous quality improvement in afterschool settings: Impact findings from the Youth
Program Quality Intervention study.* Washington, DC: The Forum for Youth Investment. Available
at: http://www.cypq.org/content/continuous-quality-improvement-afterschool-settings-impact-
findings-youth-program-quality-in

Smith, C., & Hohmann, C. (2005). Full findings from the youth PQA validation study. Ypsilanti, MI:
High/Scope Educational Research Foundation.

University of Maryland Extension (2013). Extension education theoretical framework: With criterion-
referenced assessment tools, Extension manual EM-02-2013. College Park, MD: Author. Available at:
https://extension.umd.edu/sites/default/files/_images/programs/insure/Theoretical%20Framework
%205-22-14.pdf

W. K. Kellogg Foundation (2004). Using logic models to bring together planning, evaluation, and
action: Logic model development guide. Battle Creek, MI: Author. Available at:
http://www.smartgivers.org/uploads/logicmodelguidepdf.pdf