
From Then to Now: Emerging Directions for Youth Program Evaluation

Mary E. Arnold

Oregon State University
mary.arnold@oregonstate.edu

Melissa Cater

Louisiana State University



From Then to Now: Emerging Directions for Youth Program Evaluation

Mary E. Arnold
Oregon State University

Melissa Cater
Louisiana State University

Abstract: Understanding the impact of youth development programs has been an important topic since the programs first began, and the past 25 years in particular have witnessed considerable advances in the evaluation of youth development programs. This article presents a brief history of youth development program evaluation, considering how it has changed over the years. From there, three contemporary trends related to youth program evaluation are examined: 1) a new evaluation focus, which is the emphasis on evaluating program quality; 2) organizational structures related to effective program evaluation, primarily in the area of program evaluability and evaluation capacity building; and 3) an emerging evaluation approach, involving youth in evaluating the programs that affect them. The article concludes with a call for programs to attend carefully to program implementation quality.

Introduction

Understanding the impact of youth development programs has been an important topic since the programs first began. The past 25 years in particular have witnessed considerable advances in the evaluation of youth development programs and what defines a “successful” program. The movement of evaluation from narrative accounts of program success, to counts of program participants and measures of participant satisfaction, to measures of program outcomes has taken place relatively quickly, dropping us firmly at the doorstep of the “gold standard” of program evaluation: evaluations that utilize a rigorous experimental design.

The importance of measuring program outcomes notwithstanding, recent developments in the field of youth program evaluation are setting the stage for broader, more inclusive, evaluation strategies; strategies that emphasize evaluation use and organizational learning, both of which

have been highlighted as important if evaluations are to have impact on stakeholder support, program improvement, and decision making (Fitzpatrick, Sanders, & Worthen, 2011).

In this article we will present a brief history of youth development program evaluation, and consider how it has changed over the years, exploring the developments that led to an emphasis on measuring program outcomes. As we shall see, however, not every youth development program is a good candidate for outcomes evaluation, and many youth organizations lack the resources needed to conduct rigorous outcomes evaluation. In addition, recent developments in the youth program evaluation have invited a broader understanding of the evaluative needs of youth programs. As such, the remainder of this article will consider three contemporary trends related to youth program evaluation. First we will consider a new evaluation *focus*, which is the emphasis on evaluating program quality. Second we will consider *organizational structures* related to effective program evaluation, primarily in the areas of ability of a program to be evaluated and evaluation capacity building. Finally, we will discuss the emergence of a new evaluation *approach*, involving youth in evaluating the programs that affect them.

A Brief History of Youth Program Evaluation

Publications related to youth program evaluations have flourished in the past 10 years. Where once a dearth of literature existed, today an abundance of information related to the effectiveness of youth development programs can be found. Just over 20 years ago little evaluative information on youth programs existed. Indeed, the field of program evaluation as a whole is a relatively young field of study, only now approaching the 40-year mark. Program evaluation began in the US in the early 1960s, when the first federally mandated (and funded) program evaluations got underway. The programs that underwent these early evaluations were implemented as part of the War on Poverty in the United States. Evaluators, who were largely contracted university researchers, were excited to lend their expertise to measure the effectiveness of social programs, and policy makers looked forward to programming decisions that would be based on sound evidence of a program's success. In her commentary on the impact of program evaluation during its first 25 years, Weiss (1987) reveals a rather dismal picture of the results of these early evaluations: the evaluation results did not support evidence of program success. Despite the resulting evidence, people appeared to *believe* in programs and evaluation data had little effect on program expansions or reductions. One reason for this was the recognition that social issues are complex, and the outcomes initially identified for the new Federal programs may not have been reasonable indicators of success. As Weiss (1987) points out, the yardstick used to measure success almost guaranteed failure.

Nonetheless, the results of these early evaluation efforts received important methodological critiques that began the conversation about effective and valid program evaluation; a conversation that remains strong today. First was a focus on rigor, particularly in response to the use of comparison groups rather than true experimental designs with randomly assigned control groups (Bernstein, 1975). Design critics raised the point that no evaluation can reveal valid results without a rigorous design, and attention to this would result in better evidence of program effectiveness. On the other side were those who argued the use of qualitative methods that allowed reflexive awareness and response to the "human" side of social programs, focusing on the impact of programs from the viewpoint of the program participants. This approach was deemed more useful than trying to prove outcomes that were determined a priori; outcomes that may not even be the most meaningful outcomes to measure (Patton, 1980). And

somewhere in between was the growing recognition that social programs asked a lot of people, in that participants were expected to change from their pre-intervention state to an ideal state in one step (Weiss, 1987). This expectation meant that the many important, and often critical, intermediate indicators of progress were not articulated, let alone captured as indicators of success. Nor were the human participants of programs viewed as works in progress, growing and changing in, and influenced by, the context of their lives.

Similar critiques of program evaluations are still present today; we are far from resolution. And while issues of design rigor and methodological approach remain important, the natural *developmental* influences that are at play throughout the time a youth might participate in a program complexify our ability to determine precise program factors that create success. These realities underscore the need to consider youth program evaluation a complex task, and draw into question evaluation yardsticks that do not fully consider the social and developmental contexts of youth programs.

Youth Development Program Evaluation

Youth development programs in the US began to emerge in the late 1800s and early part of the 20th century. From providing boys who “roamed the streets” of Hartford, CT in 1860 with positive alternatives (Boys & Girls Club, 2011), to helping girls become “capable and creative women” in 1910 (Camp Fire USA, 2011), to teaching rural youth about advances in agriculture through “hands-on” learning in 1902 (National 4-H Council, 2011), these early programs reflected society’s sense of social obligation to attend to the welfare and development of youth. Even in the early years, there was interest in understanding and sharing the impact of programs on youth, which often was in the form of testimonials and case studies of program participants who excelled as a result of the program.

Success stories provided heart-warming support for society’s efforts to support youth, and programs flourished in many cases because they were seen *as the right thing to do*. But changes to the economy and emerging differences in opinion about the role of society in helping youth in the 1980s ushered in a new day for youth program evaluation. As the age of accountability dawned, pressure to determine more definitively the value and impact of youth programs increased.

The formal and systematic evaluation of youth development programs did not begin until the late 1980s, when the idea of youth development as a separate program from intervention programs began to take hold. In 1989 the Carnegie Council on Adolescent Development identified five goals of successful adolescent development:

- 1) intellectually reflective;
- 2) enroute to a life of meaningful work;
- 3) good citizens;
- 4) caring and ethical; and
- 5) healthy.

While the goals were clear, clarity on what constituted a youth development program remained uncertain. Evaluations of youth programs began to take place nonetheless, with the first systematic efforts measuring program “reach,” which defined success by the number of participants in a program, and thus “proved” to funders and other stakeholders that program services were provided (Rennekamp & Engle, 2008). Measures of program reach were followed

by measures of participant satisfaction, assuming that if participants were satisfied with the program, funders would be more likely to continue funding (Rennekamp & Engle, 2008). Such indicators of success, however, did not provide evidence of program effectiveness, and it was not long before accountability expectations shifted to an emphasis on demonstrating lasting impact on program participants.

In a subsequent report, the Carnegie Council (1992) identified two sets of concerns problematic to youth development program evaluation:

- 1) lack of expertise and/or support for program evaluation, which presented an early call for evaluation capacity building; and
- 2) limitations on researchers' current approaches to evaluation, which set the stage for the development of innovative evaluation techniques.

Additional concerns were related to the lack of funding and staff allocated to outcome evaluations, even among the nation's oldest and largest youth organizations. Also of concern were evaluation designs that lacked rigor, which led to unsubstantiated claims of program success. As a result, the Carnegie Council also highlighted a need to bridge the gap between evaluators and practitioners, and perhaps most importantly, to develop consensus on what outcomes should be used to evaluate youth development programs.

In an effort to update and expand the 1992 Carnegie Council report, Roth, Brooks-Gunn, Murray and Foster (1998) attempted to synthesize youth development program evaluation. The authors searched the most relevant databases with a narrowed focus on youth development program evaluations, leaving out school-based and curricular-based programs that did not take a comprehensive youth development approach. There was such variation among the programs and the information that was provided in the evaluation that a formal meta-analysis was not possible. In the end, 15 program evaluations that had an experimental (9) or comparison group design (6) were chosen for examination.

The authors conclude that, except for a few instances, little improvement to the state of youth program evaluation development had occurred since the 1992 Carnegie Report. They conclude that the lack of quality evaluations could be related to the newness of the youth development framework and if true, then improvement should occur and be evidenced by increased literature related to youth program evaluation. This early paucity of rigorous program evaluations is important to note, as it sets a baseline for understanding the development of youth program evaluation in the subsequent 13 years.

Despite the early lack of high quality program evaluations, indicators of success for the youth development framework were beginning to emerge. The strongest themes for these indicators were:

- 1) the presence of adults who fostered skill, community building and hope for youth;
- 2) youth who were seen as resources to be developed rather than problems to be fixed; and
- 3) programs that created spaces of belonging where youth feel safe, cared for, and empowered.

The particular activities of the programs were not as important as the program's ability to create an atmosphere for active participation and opportunities for challenge and growth.

By the early part of this century a better articulation on youth program structure and outcomes developed. As part of this, Eccles and Gootman (2002) concluded that all youth programs should undergo evaluation, but the goals, and thus the design, for evaluation will differ from program to program. The authors concur that very few high-quality experimental program evaluations had been conducted to determine the impact of programs on youth. While they cite many possible reasons for this, the primary factor is that such evaluations take time, money and knowledge resources, things that most youth serving organizations do not have. The authors acknowledge that comprehensive experimental designs are still critically important, but such designs need to be coupled with evaluations of program implementation in order to understand better the factors behind the effects found through experimental designs. Furthermore, only programs that meet certain criteria should even consider experimental designs. These criteria include evaluating only program components that are common to many youth programs and limiting such evaluations to established national organizations with local affiliates. Although non-experimental designs reveal little about program effect, they are useful for assessing program implementation and identifying patterns of effective practice. According to Eccles and Gootman (2002) candidates for non-experimental designs include programs that are quite broad, relatively immature, when the goal of the evaluation is to assess fidelity and program implementation, or when the program staff is responsible for conducting the evaluation.

Emerging Trends in Youth Development Program Evaluation

The importance of comprehensive, rigorous, experimental studies notwithstanding, many youth serving programs lack the resources to conduct comprehensive studies and even if they could, they may not provide the most useful information to the program. Recently, three important trends in youth development programming have begun to emerge that help broaden the way we think about youth program evaluation methods and use. The first, which represents a change in evaluation *focus*, involves an emphasis on evaluating program quality. The second underscores the importance of *organizational* support for evaluation through evaluation capacity building. And the third reflects a shift in evaluation *approach* by involving youth in evaluating the programs that serve them.

Measuring Program Quality as a Critical Factor in Youth Program Evaluation

Recently, researchers and program evaluators alike began to question why some programs were achieving targeted outcomes while others were not. Momentum built around discovering the reason for the lack of consistency in achieving youth-level outcomes, and the phrase *program quality* began to emerge. Early definitions of program quality had a global focus on attaining high standards of practice and achieving targeted outcomes (Pittman, Tolman, & Yohalem, 2005). With the entrance of the National Research Council's (Eccles & Gootman, 2002) eight features of positive developmental settings, interest at all levels of practice, research, and policy swung toward describing what was happening within programs as a way of adding explanatory power to the achievement, or lack thereof, of targeted youth outcomes. This list of program features built upon the themes emerging from developmental theory, empirical research in educational and family settings, and early youth program evaluations.

The race to "scale up" program implementation led some researchers to investigate more closely the association between program implementation and youth outcomes. It became increasingly apparent that the transition from a carefully controlled research study to a practical, real-world youth program resulted in a breakdown of the quality of program delivery (Gerstenblith, et al., 2005). Many programs were sacrificing key elements related to staffing

patterns and program features that promoted youth engagement in their quest to offer services to a larger audience.

With evidence of the link between program quality and youth outcomes mounting, attention turned to considering which specific program features were most important in achieving outcomes. Building from Eccles and Gootman's (2002) features of positive developmental settings, many programs developed their own list of high quality program content, structures, and processes. Program practices that have been almost universally adopted include creating spaces of physical and psychological safety, building supportive relationships, delivering effective programming, and providing opportunities for youth engagement (Granger, Durlak, Yohalem, & Reisner, 2007; Grossman, Goldsmith, Sheldon, & Arbreton, 2009; Hirsch, Mekinda, & Stawicki, 2010; National Institute on Out-of-School Time, 2000; Siaca, 2010). A less commonly selected program characteristic was cultural sensitivity. Yet emerging evidence supports the significance of this program feature in the achievement of youth outcomes in programs with diverse audiences (Pierce, Bolt, & Vandell, 2010; Riggs, Bohnert, Guzman, & Davidson, 2010). As support for the focus on program practices and quality has grown, the definition of program quality has drifted away from quality as a global concept and instead has concentrated on quality at the point of service with convergence around the importance of the interaction between program content, staff practices and youth experiences (Hirsch et al., 2010; Smith, Peck, Denault, Blazevski, & Akiva, 2010; Yohalem, Granger, & Pittman, 2009).

The interest in evaluating program quality has expanded quite rapidly over the last 12 years. We have progressed through stages of simply understanding what was meant by program quality, questioning why it was important, and determining how to measure it, to considering the obligation of accountability for both program quality and youth outcomes (Yohalem, Granger, & Pittman, 2009). As the program quality movement has grown, the question of minimum levels of quality and program improvement has developed simultaneously. The idea that "programs are only as good as their implementation" (Hirsch, et al., 2010, p. 450) points to a need to adhere to the evidence-based program delivery protocol and alludes to the two entry points for evaluation in the program delivery process. First, as youth programs shift from a research environment to a practical program, it is common for staff to make adjustments in program delivery. These implementation changes may be simple scheduling alterations that have little impact on program quality and youth outcomes or modifications such as reducing the number of staff members that may significantly affect the program's ability to achieve quality standards. Implementation evaluation monitors the fidelity of program delivery and is used in conjunction with short-term youth outcome data to determine if program changes are negatively affecting youth outcomes. Second, as programs age and staffs change, evaluation of both program delivery and the quality of that delivery become critical to the successful maturing of the program.

As policymaker and funder interest in program quality increases, the pressure on programs to respond will also increase. Youth development programs may begin to position themselves for this increased accountability by creating systems that track both point of service quality and youth outcomes (Granger, Durlak, Yohalem, & Reisner, 2007) and the relationship between the two. Practitioners and researchers alike are cognizant that the youth development field could quickly follow the path of the education (teaching to the test) and prevention (serving less needy youth) fields if expectations become too stringent (Yohalem, Granger, & Pittman, 2009). The lessons learned from our counterparts' experiences serve as a valuable foreshadowing of the challenges the youth development field may face if the accountability expectations are set too high. Eccles and Gootman (2002) bring focus to the accountability discussion through their

entreaty to shape the scope and rigor of evaluations to the goals and resources of individual programs. This serves as a reminder that not every program can or should be evaluated.

Evaluation Capacity Building through Youth Participatory Evaluation

The field of program evaluation has grown exponentially in the past 40 years. So much so, that Hallie Preskill, in her 2007 presidential address to the American Evaluation Association, claimed that the field of evaluation had arrived at a "Tipping Point", a *liminal* place where something wholly new was about to emerge (Preskill, 2008). Referring to what she termed "evaluation's second act," Preskill emphasized the critical importance of building evaluation capacity of people and their organizations to create cultures of evaluation to think evaluatively, engage in evaluation practice, and use evaluation findings. The movement toward evaluation capacity within organizations is especially important to explore in the field of youth development, particularly given the fact that so many youth programs do not meet the criteria for comprehensive experimental evaluations outlined by Eccles and Gootman (2002). Many youth organizations are struggling to find ways to develop internal evaluation capacity, often because of the expectations of external funders, both large and small, but also because the organizations *want* to know about the impact of their programs on the youth they serve.

Evaluation capacity building (ECB) is defined as an *intentional* process to create and sustain an organizational culture that routinely conducts evaluations and uses the evaluation results (Comptom, Bazierman, & Stockdill, 2002). One of the most important aspects of ECB is the emphasis on organizational learning and development, as it is now understood that building individual evaluation capacity alone will not do enough to create quality evaluation practice in organizations (Preskill & Boyle, 2008). As Taylor-Powell and Boyd (2008) point out, ECB can be messy business, especially in complex organizations. Building evaluation capacity and doing evaluations are not the same thing and the two roles are often confused, especially when organizational understanding of, and support for, evaluation is lagging.

Taylor-Powell and Boyd (2008) outline a three-part framework for ECB that includes 1) professional development; 2) resources and support; and 3) organizational environment. This framework is useful for understanding that professional development (i.e. individual capacity) alone is not enough. Careful attention must be paid to the resources and organizational culture if evaluation capacity is to be developed and sustained. As more and more youth organizations seek to build evaluation capacity, certain important and interesting elements are emerging. In particular we focus on the need for providing "just in time" evaluation training for youth organizations, and the practice of involving youth in the evaluation of the programs that serve them.

When applying ECB efforts to youth serving organizations, one of the first complications that arises is the need to build capacity *and* conduct evaluations at the same time. Unlike other professional development opportunities that typically build on a previously established professional foundation, many professionals in youth-serving organizations have little training in program evaluation. Youth programs are often driven to seek training because of external and immediate expectations for evaluation data. In these situations youth programs do not have the luxury of learning all they need to know before beginning an evaluation. Arnold (2006) proposed a tested framework for building evaluation capacity with 4-H youth development educators. This framework consisted of four strategies: 1) using logic models for articulating program plans and theory; 2) providing one-on-one evaluation assistance; 3) facilitating small-team collaborative evaluations; and 4) conducting larger-scale evaluations. In this instance the author was an internal evaluator working side by side with program staff to build evaluation

capacity while conducting internal evaluations at the same time. While the framework Arnold proposes was effective, most youth organizations do not have an internal evaluator to do this work. Others have proposed frameworks that are collaborative efforts between external evaluators and program staff that have demonstrated ECB effectiveness (Garcia-Itiarte, Suarez, Balcazar, Taylor-Ritzler, & Luna, 2010; Huffman, Thomas, & Lawrenz, 2008).

Although collaborative and internal ECB strategies show promise, the overall need for evaluation capacity building remains largely unaddressed. We suspect that the evaluative needs of youth development programs far outweigh the professional evaluation capacity and resources to meet those needs. However, in the youth development arena a new approach to program evaluation is gaining considerable momentum, and that is engaging youth in participatory evaluations of the programs that serve them. This approach, often called Youth Participatory Evaluation (YPE), has a double impact in that programs gain valuable evaluation data and youth gain developmentally. Youth participatory evaluation may well be an example of development in the limited approaches for evaluating youth programs identified in the 1992 Carnegie Council report.

Youth Participatory Evaluation

Participatory evaluation, with its emphasis on the *practical* use of evaluation findings and the *transformative* effect it can have on program participants (Cousins & Whitmore, 1998), has attracted significant interest from evaluators seeking a more holistic approach to program evaluation. In addition, involving youth in participatory evaluation had become increasingly common in the past eight years (Arnold, Dolenc, & Wells, 2008; Camino, Zeldin, Mook, & O'Conner, 2004; Checkoway & Gutierrez, 2007; Checkoway & Richards-Schuster, 2006; Chen, Weiss, & Johnston-Nicholson, 2010; Delgado, 2006; Fetterman, 2003; London, Zimmerman, & Erbsstein, 2003; Sabo, 2003, Sabo Flores, 2008). Engaging youth in the evaluation of the programs that affect them has powerful potential, while at the same time facilitating and demonstrating the values and outcomes of positive youth development programs.

A recent youth participatory evaluation conducted by Girls Incorporated (Girls Inc.) (Chen, Weiss, & Johnston-Nicholson, 2010) highlights many of the converging factors that support the potential of this approach. In this evaluation, girls ages 12-18 formed research teams to evaluate the effectiveness of the Girls Inc. program. The evaluation questions focused on how the program helps girls achieve the program's stated goals (e. g. inspiring girls to be strong, smart, and bold) as well as how the program could better meet the needs of girls and their communities. Two key forces provided the impetus for the study: 1) the desire to involve girls in diverse leadership and advocacy roles; and 2) the increasing demand for "measurable and convincing evidence" of the positive impact of the program. Although the project was deemed successful, considerable support and resources contributed to the success. The national organization provided financial support as well as research and evaluation expertise. Each site was trained using a common curriculum and ongoing technical assistance and support was provided to the local affiliates. The success of the project did not "just happen" but was the result of careful planning, use of evidence-based practices, and adequate training and support. It is important to keep in mind that evaluations employing more traditional designs and methods also do not "just happen" but require similar investment of time and resources.

When considering ECB through youth participatory evaluation, several strengths come to mind. First, unlike methods that build staff capacity to conduct evaluations, building youth capacity along with staff automatically secures the capacity at a larger organizational level. The youth themselves become invested in the evaluation, thus increasing the likelihood of a positive

evaluation culture within the program. Second, involving youth in evaluation becomes the program itself. As we know, evaluation efforts can often be viewed as “add-on” activities that need to be done in addition to programming. With YPE, the evaluation becomes the program method itself, employing well-established program elements such as youth-adult partnerships (Camino, 2005; Zeldin, Larson, Camino, & O’Conner, 2005) that contribute to the positive development of the youth participants while at the same time conducting the evaluation. Going back to Eccles and Gootman’s (2002) summary of the usefulness of non-experimental designs for assessing program implementation and identifying patterns of effective practice, YPE has strong potential for gathering meaningful and reliable data as youth are often more willing to open up and share their feelings with other youth than adult researchers.

Conclusion

We opened this paper with the goal of providing a timely and useful lens through which to view the evaluation of youth development programs. The field of positive youth development, with the particular definitions and criteria that define it, has matured considerably in the past 20 years. Likewise, issues related to the best practices for the evaluation of youth programs have grown in tandem. The call from the Carnegie Council to develop expertise in program evaluation and to find innovative new methods for conducting valid evaluations remains a driver in youth program evaluation today.

There is no question that comprehensive, random experimental evaluation designs remain the “gold standard” in the minds of all who struggle to define what is meant by acceptable evidence for program effectiveness. This is especially true for providing evidence to garner political and financial support for programs. Related, and of equal concern for many, is the articulation of program outcomes and valid methods for determining a program’s effect on those outcomes. Unfortunately, the emphasis on rigorous outcome evaluation can be a barrier to the development of additional evaluation strategies that are more appropriate, meaningful and useful for some youth programs. It is highly unlikely that these concerns will be fully addressed as we move forward; rather they will assume a perennial role in the debate over what constitutes acceptable evidence.

Meanwhile, while the debate rages on, youth programs large and small, operated by staff with scarce resources, and even less evaluative experience will continue valiantly to make a difference in the lives of the youth with whom they work. These practitioners will bear steady witness to their own success, often through the narrative stories of the youth who blossom in their programs.

As practicing evaluators, our hope is that this article encourages youth development practitioners to attend to program quality and implementation and the resulting link to program outcomes. Without sound program implementation, an evaluation of outcomes is meaningless. Likewise, we hope for the development of better evaluation capacity building frameworks, and that practitioners will begin to involve youth as evaluators of the program that affect them.

References

Arnold, M.E. (2006). Developing evaluation capacity in Extension 4-H field faculty: A framework for success. *American Journal of Evaluation*, 27(2), 257-269.

- Arnold, M.E., Dolenc, B., & Wells, E.E. (2008). Youth community engagement: A recipe for success. *Journal of Community Engagement and Scholarship*, 1, 56-65.
- Bernstein, R.N. (1975). Validity issues in evaluative research: An overview. *Sociological Methods and Research*, 4, 3-12.
- Boys & Girls Club. (2011). *Our history*. Retrieved May 5, 2011, from <http://www.bgca.org/whoweare/Pages/History.aspx>.
- Camino, L., Zeldin, S., Mook, C., & O'Conner, C. (2004). *A practical guide for program assessment and action planning*. Madison, WI: University of Wisconsin.
- Camino, L. (2005). Pitfalls and promising practices of youth-adult partnerships: An evaluator's reflections. *American Journal of Community Psychology*, 33, 75-85.
- Camp Fire USA. (2011). *About camp fire*. Retrieved May 5, 2011 <http://www.campfireusa.org/History.aspx>.
- Carnegie Council on Youth Development. (1989). *Turning points: Preparing American youth for the 21st century*. Washington, DC: Author.
- Carnegie Council on Youth Development. (1992). *Consultation on evaluation of youth development programs: Report on the Meeting*. Washington, DC: Author.
- Checkoway, B., & Richards-Schuster, K. (2006). *Participatory evaluation with young people*. Ann Arbor, MI: University of Michigan School of Social Work.
- Checkoway, B., & Gutierrez, L.M. (2007). *Youth Participation and Community Change*. Thousand Oaks, CA: Sage.
- Chen, P., Weiss, F.L., Johnston-Nicholson, H. (2010). Girls study Girls Inc.: Engaging girls in evaluation through participatory action research. *American Journal of Community Psychology*, 46, 228-237.
- Compton, D.W., Baiserman, M., & Stockdill, S.D. (Eds.). (2002). The art, craft, and science of evaluation capacity building. *New Directions for Evaluation*, 93, 7-25.
- Cousins, J.B., & Whitmore, E. (1998). Framing participatory evaluation. E. Whitmore (Ed.). *New Directions for Evaluation: Vol. 80. Understanding and practicing participatory evaluation* (pp. 5-24). San Francisco: Jossey-Bass.
- Delgado, M. (2006). *Designs and methods for youth-led research*. Thousand Oaks, CA: Sage.
- Eccles, J., & Gootman, J. (Eds.). (2002). *Community Programs to Promote Youth Development*. Washington, DC: National Academy Press.
- Fetterman, D. (2003). Youth and evaluation: Empowered social change agents. In K. Sabo (Ed.). *New Directions for Evaluation: Vol. 98. Youth participatory evaluation: A field in the making* (pp 87-92). San Francisco: Jossey-Bass.
- Fitzpatrick, J.L., Sanders, J.R., & Worthen, B.R. (2011). *Program evaluation: Alternative approaches and practical guidelines* (4th Ed.). Upper Saddle River, NJ: Pearson Education.

Garcia-Iriarte, E., Suarez-Balcazar, Y., Taylor-Ritzler, T., & Luna, M. (2010). A catalyst-for-change approach to evaluation capacity building. *American Journal of Evaluation, 21*, 1-15.

Gerstenblith, S., Soule, D., Gottfredson, D., Lu, S., Kellstrom, M., Womer, S., & Bryner, S. (2005). After-school programs, antisocial behavior, and positive youth development: An exploration of the relationship between program implementation and changes in youth behavior. In J. Mahoney, R. Larson, & J. Eccles (Eds.), *Organized activities as contexts of development: Extracurricular activities, afterschool and community programs*. Mahwah, NJ: Lawrence Erlbaum Associates.

Granger, R., Durlak, J., Yohalem, N., & Reisner, E. (April, 2007). *Improving after-school program quality*. New York, NY: William T. Grant Foundation.

Grossman, J., Goldsmith, J., Sheldon, J. & Arbreton, A. (2009). Assessing afterschool settings. *New Directions for Youth Development, 121*, 89-108.

Hirsch, B., Mekinda, M., & Stawicki, J. (2010). More than attendance: The importance of after-school program quality. *American Journal of Community Psychology, 45*, 447-452.

Huffman, D., Thomas, K., & Lawrenz, F. (2008). A collaborative immersion approach to evaluation capacity building. *American Journal of Evaluation, 29*, 358-368.

London, J.K., Zimmerman, K., & Erbstein, N. (2003). Youth-led research and evaluation: Tools for youth, organizational and community development. In K. Sabo (Ed.). *New Directions for Evaluation: Vol. 98. Youth participatory evaluation: A field in the making* (pp 33-46). San Francisco: Jossey-Bass.

National 4-H Council. (2011). *The birth of 4-H programs*. Retrieved May 5, 2011, from <http://www.4-h.org/about/4-h-history/>.

National Institute on Out-of-School Time. (June, 2000). *Making an impact on out-of-school time: A guide for corporation for National Service programs engaged in after school, summer, and weekend activities for youth people*. Wellesley, MA: Author.

Patton, M.Q. (1980). *Qualitative Evaluation Methods*. Thousand Oaks, CA: Sage.

Pierce, K., Bolt, D., & Vandell, D. (2010). Specific features of after-school program quality: Associations with children's functioning in middle childhood. *American Journal of Community Psychology, 45*(3&4), 381-393.

Pittman, K., Tolman, J., & Yohalem, N. (2005). Developing a comprehensive agenda for the out-of-school hours: Lessons and challenges across cities. In J. Mahoney, R. Larson, & J. Eccles (Eds.), *Organized activities as contexts of development: Extracurricular activities, afterschool and community programs*. Mahwah, NJ: Lawrence Erlbaum Associates.

Preskill, H. (2008). Evaluation's second act: A spotlight on learning. *American Journal of Evaluation, 29*, 127-138.

Preskill, H., & Boyle, S. (2008). A multidisciplinary model of evaluation capacity building. *American Journal of Evaluation, 29*, 443-459.

Rennekamp, R.A., & Engle, M. (2008). A case study in organizational change: Evaluation in Cooperative Extension. In M.T. Braverman, M. Engle, M.E. Arnold, & R.A. Rennekamp (Ed.). *New Directions for Evaluation: Vol. 120. Program evaluation in a complex organizational system: Lessons from Cooperative Extension* (pp 15-26). San Francisco: Jossey-Bass.

Riggs, N., Bohnert, A., Guzman, M., & Davidson, D. (2010). Examining the potential of community-based after-school programs for Latino youth. *American Journal of Community Psychology, 45*, 417-429.

Roth, J., Brooks-Gunn, J., Murray, L. & Foster, W. (1998). Promoting healthy adolescents: Synthesis of youth development program evaluations. *Journal of Research on Adolescence, 8*(4), 423-459.

Sabo, K. (2003). Editors notes. In K. Sabo (Ed.). *New Directions for Evaluation: Vol. 98. Youth participatory evaluation: A field in the making* (pp 1-11). San Francisco: Jossey-Bass.

Sabo Flores, K. (2008). *Youth participatory evaluation: Strategies for engaging young people*. San Francisco, CA: Jossey Bass.

Siaca, J. (2010). High-impact afterschool for all. *Afterschool Matters, 11*, 1-6.

Smith, C., Peck, S., Denault, A., Blazeovski, J., & Akiva, T. (2010). Quality at the point of service: Profiles of practice in after-school settings. *American Journal of Community Psychology, 45*, 358-369.

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

Weiss, C.H. (1987). Evaluating social programs: What have we learned? *Society, 25*(1), 40-45.

Yohalem, N., Granger, R., & Pittman, K. (2009). The quest for quality: Recent developments and future directions for the out-of-school-time field. *New Directions for Youth Development, 121*, 129-140.

Zeldin, R., Larson, R., Camino, L. & O'Conner, C. (2005). Intergenerational relationships and partnerships in community programs: Purpose, practice, and direction for research. *Journal of Community Psychology, 33*, 1-10.