

Fall 10-20-2010

Evaluating Programs Serving High-Risk Populations: Challenges and Recommendations

Gavrielle Levine

Long Island University, glevine@liu.edu

Peter Swerdzewski

pswerdz@me.com

Follow this and additional works at: https://opencommons.uconn.edu/nera_2010

 Part of the [Education Commons](#)

Recommended Citation

Levine, Gavrielle and Swerdzewski, Peter, "Evaluating Programs Serving High-Risk Populations: Challenges and Recommendations" (2010). *NERA Conference Proceedings 2010*. 23.
https://opencommons.uconn.edu/nera_2010/23

Evaluating Programs Serving High-Risk Populations: Challenges and Recommendations

Abstract

Grant-supported programs, particularly those sponsored with government funds, are required to include a rigorous evaluation as part of the proposal and project implementation to demonstrate program effectiveness. Ironically, programs that serve high-risk populations are often at an inherent disadvantage with regard to conducting rigorous program evaluations since this unique population does not lend itself to accepted program evaluation standards and practices. Consequently, these programs often do not receive the funding needed to provide useful programs for high-risk populations. The New York City-based RESOLVE abstinence education program will be used as a running example to discuss challenges and recommendations that are commonly encountered when evaluating programs that target high-risk youth. Examining the challenges that high-risk populations present to evaluators can lead to the development of new evaluation protocols that are more appropriate to this group of individuals.

Grant-supported programs, particularly those sponsored with government funds, are required to include a rigorous evaluation as part of their proposal and implementation process in order for the program to demonstrate it is effective in meeting its intended outcomes. Ironically, programs that serve high-risk populations are often at an inherent disadvantage with regard to conducting rigorous program evaluations because this unique population does not lend itself to accepted program evaluation standards and practices. Consequently, these programs often do not receive the funding needed to provide useful programs for these important high-risk populations.

In conducting evaluations of programs that target high-risk populations, evaluators frequently find themselves asking, “What must I do to amend common evaluation practices to fit these unique students?” and “How much can I diverge from common evaluation standards to meet the needs of unique populations without the threat of stakeholders discounting the results for my evaluation?” These and other critical questions are yet unanswered in the program evaluation field; however, as the field evolves it has become necessary to start addressing these questions to ensure the needs of high-risk populations are met. The New York City-based RESOLVE abstinence education program will be used as a running example in this paper to discuss challenges and recommendations that are commonly encountered when evaluating programs that target high-risk youth. RESOLVE’s target population includes youth who live in foster homes as well as many youth who currently live in residential facilities specifically created for individuals who struggle in traditional foster family settings.

By taking a strategic look at common challenges inherent to the evaluation of programs involving high-risk youth, this research will ideally inform future practices in the domain. Because the challenges and recommendations extend from a practice example (the RESOLVE

program) rather than an abstract theoretical illustration, others in the field may find this work useful and can build from this work, thereby contributing to the burgeoning literature on the evaluation of high-risk youth.

Challenges Encountered When Evaluating a Program for High-Risk Youth

The techniques and precepts of the educational evaluation community (e.g., Sanders, 1994) have relied heavily on studies and policies derived from traditional youth populations (For a discussion of the evolution of educational program evaluation, see W.K. Kellogg Foundation, 1998.); however, not all evaluation techniques are appropriate for all youth populations (Rossi, Lipsey, and Freeman, 2004). The population of youth that serve as the focus of this paper—high-risk youth—present unique challenges that preclude the use of many commonly accepted evaluation practices.

Historically, the field of program evaluation has sought to keep pace with changing demographics and evolving needs (Smith, 1981). Although early demand for rigorous program evaluation focused on mainstream populations (e.g., Kirkpatrick 1959a, 1959b, 1960a, 1960b), more contemporary research has focused on developing custom evaluation solutions for unique high-risk populations including youth who exhibit violence behavior (Parker and Hudley, 2006), youth who are subject to mental health services (Bickerman and Rog, 1992), and individuals who are homeless (Rog, 1991). Despite these advances in the field, there is still relatively little research that clearly defines how to effectively evaluate programs that target various subpopulations of high-risk youth, although there are growing niches within the field that are informing current and planned work in the area. For example, the work of J.B. Jemmott has provided a strong foundation upon which abstinence education research has been building over

the past few years (see Jemmott and Fry, 2002 for a review of some foundational findings), which has in turn informed the evaluation of the abstinence education program discussed later in this paper.

When amending common protocol used in program evaluation, it is important that guiding principles be kept in mind. For example, revised evaluation protocol should abide as closely as possible to the *Program Evaluation Standards, 2nd Ed.* (Sanders, 1994) created by the Joint Committee on Standards for Educational Evaluation. Likewise, amended protocol should take into consideration the ramifications of social constructs such as race (Madison, 1992), culture (Thompson-Robinson, Hopson, and SenGupta, 2004), and language (Hopson, 2000). Furthermore, as one would expect, values and ethics are key issues that must be considered when amending common evaluation protocol to accommodate unique populations such as high-risk youth (Perloff and Perfloff, 1980).

One particularly helpful framework that can help guide program evaluators in determining what amendments are appropriate to make to commonly-accepted program evaluation protocol is Dencombe's 10-point "Ground Rules for Good Research" (2002). Although written for academic research, this framework is highly-accessible to program evaluators due to its clarity and focus on making strong, valid inferences from programmatic data. The 10-point framework is as follows (p. 4):

To qualify as social research an investigation needs to:

- Have clearly stated aims, that are
- related to existing knowledge and needs, and that are
- investigated within limitations imposed through time, money and opportunity.

Research needs to:

- contribute something new to knowledge, using
- precise and valid data,
- collected and used in a justifiable way,
- to produce findings from which generalizations can be made.

The researcher needs to adopt an attitude and approach that is:

- open-minded and self-reflective,
- cognizant of the rights and interests of participants, and
- cautious about claims based on the findings.

Taken together, there currently exists a sufficient body of research that program evaluators can reference when identifying factors that must be considered when attempting to amend common evaluation protocol to meet the needs of unique populations. Furthermore, existing frameworks such as that proposed by Dencombe can provide a unifying model when attempting to organize and assess the impact of possible changes to common program evaluation protocol. It is through the effective use of these factors and frameworks that program evaluators will be able to successfully amend program evaluation protocol to meet the unique needs of subpopulations.

Examining Issues Surrounding High-Risk Youth Using the RESOLVE Program as a Model

The program evaluated in this study of high-risk youth is known as the RESOLVE program. RESOLVE is a successful program that was developed by the Jewish Child Care Association (JCCA) and supported with a Community-Based Abstinence Education (CBAE) grant, administered by the Department of Health and Human Services Administration for

Children and Families, to provide abstinence education and encourage healthy lifestyles to underserved populations of high-risk youth in foster care and community-based organizations.

About the RESOLVE Program

RESOLVE's 16-hour eight-module curriculum teaches strategies to promote abstinence among high-risk youth, along with skills to support this healthy lifestyle choice. Implementation of the curriculum was designed to encourage youth to develop healthy relationships both in the short- and long-term, while emphasizing the importance of setting long-term life goals and highlighting the key role of abstinence in accomplishing those goals by avoiding early pregnancy and sexually transmitted diseases. Training in assertiveness and social skills was combined with those specific tools and strategies necessary to implement abstinence and other healthy lifestyle choices. Experiential and interactive activities, designed to enhance goal setting, refusal skills, and self-esteem, engaged these hard-to-reach youth.

The RESOLVE program also offered training in abstinence education and healthy lifestyles to foster parents and parents as well as youth workers in foster care and community settings who work with these high-risk youth.

Evaluation of the RESOLVE Program

The evaluation includes data on both implementation and outcomes assessment for three years of the RESOLVE program. It examines program delivery, including the number of youth served across the various program settings and progress toward meeting the program's participation targets. In addition, the outcomes assessment provides data on program-year outcomes, including changes in content knowledge, attitudes and intentions, and self-esteem for

participating youth. A pre-test/post-test within subjects design is used to test the following hypotheses:

- Participating in the RESOLVE program will increase youths' knowledge about the benefits of remaining abstinent.
- Participating in the RESOLVE program will increase youths' knowledge of refusal skills, and the benefits of healthy life choices.
- Participating in the RESOLVE program will increase participants' self-esteem.

Although the purpose of this paper is to discuss the challenges and possible solutions inherent to evaluating at-risk youth, it is both interesting and important to note that the evaluation of the RESOLVE program demonstrated that youth experienced statistically significant increases in knowledge and self-esteem in conjunction with attending the program's curriculum-based sessions.

Recommended Solutions for Evaluating a Program for High-Risk Youth

Evaluating programs that involve high-risk youth is clearly a challenge. High-risk youth tend to have different backgrounds and often engage in different behaviors than their more traditional youth counterparts. As such, common evaluation practices do not always apply to programs that involve high-risk youth. Researchers must engage in the creative process to develop evaluation plans that are tailored to the unique life situations of high-risk youth, yet are also accepted as rigorous evaluation processes from which one can make valid inferences about program effectiveness.

The Strategic Steps Involved in Understanding the Impact of High-Risk Youth on Educational Program Evaluations

In assessing the RESOLVE program to determine suitable amendments to commonly accepted evaluation protocol, a number of challenges emerged related to these high-risk youth. In order to move forward with the critical evaluation of this very important program, it was important to methodically examine these challenges. The deliberate process of examining these challenges, which in turn informed the changes that were realized to the evaluation protocol, was a multi-step intentional process.

Step one: Understanding the youth group. First, the evaluation team partnered with the program management team to identify commonalities among the youth who were participants in the program. As noted above, it was found that this specific cadre of high-risk youth suffered from higher-than-average instances of physical, mental, and sexual abuse. Similarly, these youth were often raised in homes in which alcoholism or drug abuse was a persistent issue. In all cases, the population of youth addressed in this study (known as the RESOLVE foster care youth) were not living with their parents at the time of their participation in the RESOLVE program. Most youth were subject to court-ordered placement in foster homes or residential foster facilities. Some youth were in detention centers (i.e., juvenile hall).

Step two: Identifying instances in which the youth group does not lend itself to traditional program evaluation techniques. Next, the evaluation team sought to identify those instances in which traditionally accepted program evaluation techniques were not appropriate for implementation due to the unique issues surrounding the non-traditional populations. This process required a thorough understanding of accepted practices in program evaluation, as well as a deep recognition of what processes and policies were appropriate for this unique population

of youth participants. For example, through this discovery process it became apparent that youth in the RESOLVE program required developmentally appropriate assessments that were outside the ambit of those assessments normally administered to mainstream youth. Similarly, in considering the ethics surrounding evaluation of this group of youth it became apparent that a control group was not an ethically viable option because it was understood that all youth could reasonably benefit from the curriculum-based intervention; thus, denying the treatment from a group of youth for the purpose of evaluation was considered unethical.

Step three: Engage in intensive creation of amended program evaluation protocol to accommodate these high-risk youth. The third step of the process was to amend the traditional evaluation protocols that are often used in the evaluation of youth-focused education programs to accommodate the unique high-risk population. This effort required careful consideration of the requirements of the program's evaluation to ensure that amendments to normally accepted protocol would still yield data from which program evaluators could make strong inferences about the degree to which youth learned and grew as a result of the curriculum-based treatment program.

Step four: Implement revised program evaluation protocol and assess effectiveness of the new protocol. Finally, the revised program evaluation protocol were implemented within the operational treatment program. In some instances this was done on a pilot-study basis, whereas in other cases the revised protocol was introduced across the entire program. Importantly, although all protocol were vetted ahead of implementation, the revised protocol were also closely

examined for their effectiveness once implemented to ensure they were appropriate in the operational setting.

Challenges and Recommendations Identified for the RESOLVE Program

Through the strategic discovery process outlined above, it became clear that the unique experiences of high-risk youth presented specific challenges to the RESOLVE program, and as such it was critical that the RESOLVE program's protocol be amended to accommodate these challenges to contemporarily-held beliefs about what constitutes accepted protocol for the rigorous evaluation of educational programs. These issues included the following: (1) the inability of the RESOLVE program to accommodate a control group with random assignment, (2) the transient nature of the population, (3) sensitivity issues surrounding the high-risk population, and (4) the creation of materials that were developmentally-appropriate for this group of high-risk youth. Each of these challenges that were experienced by the RESOLVE abstinence education program evaluation team is discussed next.

Inability to use a control group with a randomized design. One significant challenge confronting the evaluators of the RESOLVE program was the inability to implement a rigorous experimental design (i.e., a design in which participants are randomly assigned to a treatment or a control group). The last decade has seen the scientific method increasingly influence program evaluation designs as stakeholders have called for evaluations that yield stronger inferences about the growth that can be attributed to a given program, which in turn has made experimental designs the gold-standard for program evaluation (Shadish, Cook, and Campbell, 2002). Experimental designs are challenging to implement, and research has unsurprisingly

demonstrated that evaluators have been slow to fully adhere to the new demands for experimental designs in program evaluations (Christie and Fleischer, 2010) despite the upswing in the prevalence with which experimental designs are required by program stakeholders (e.g., Brandon et al., 2010).

The RESOLVE program is one such program in which the evaluation team found it challenging to implement a traditional experimental design. High-risk youth in the RESOLVE program live in foster homes or residential facilities that make randomized delivery of services administratively impractical and, in some cases, unethical. In most cases, the implementation of a traditional experimental design was administratively impractical due to the constraints imposed by the structure of the foster homes and residential facilities in which the youth participants lived, as well as the extremely limited resources available to implement such a design. In short, although the implementation of an experimental design was perceived as important to nearly all stakeholders, it was seen as more important to ensure that the basic needs of the youth involved in the foster homes and residential facilities were met *first*, which in turn precluded implementation of an experimental design across all program participants. Likewise, because the RESOLVE program was viewed as meeting a critical need for these youth, many stakeholders argued that providing the program to only a subset of the population (i.e., those *not* assigned to the control group) would be unethical. Said another way, because the program could potentially prevent a student from contracting a serious disease, becoming pregnant, and/or being placed in a situation that could lead to abuse, program stakeholders reasoned that all youth should have the opportunity to benefit from the program.

This challenge was addressed through the implementation of a unique and seldom-used research design. Because stakeholders required that all youth receive the treatment program, a

control group in which *no* youth received the treatment was not an option; therefore, program evaluators implemented a staggered randomized design that allowed all students to participate in the treatment while they essentially serve as their own control group (Trochim, 2006). In this design, youth were randomly assigned to one of two groups. Youth in both groups took the pretest, then youth in the first group participated in the treatment program while youth in the second group had no exposure to the program and thus served as a temporary control group to youth in the first group. Both groups were then tested—the first group taking what equated to a posttest and the second group taking what equated to a second pretest. The second group then participated in the treatment (thereby allowing all youth to experience the program) and both groups were again tested at the conclusion of the second group's participation in the program.

The randomized staggered groups design as implemented for the evaluation of the RESOLVE program is not common in program evaluation; however, it allows for much stronger inferences than a design in which there are fewer controls and thus limited inferences about the effectiveness of the program. The unique design was an example of a creative approach to addressing the challenges that arise from serving a high-risk youth population such as the one served by the RESOLVE program. A valuable byproduct of the use of this stronger design is the increased chance that scholarly journals will accept the design as sufficiently rigorous for publication, which allows other researchers the opportunity to benefit from the findings of the RESOLVE program.

Transient nature of the population. Yet another challenge that confronted the evaluators of the RESOLVE program was the transient nature of the target youth population. Traditional evaluation techniques tend to have strict requirements about subject participation and attrition to

make valid inferences from evaluation data. A traditional evaluation would therefore not accommodate the degree to which youth enter and exit the RESOLVE study due to unique life circumstances. Frequently, foster care youth had required hospital visits, family visits, and sometimes simply were absent without leave which prevented them from attending curriculum sessions and completing evaluation measures consistently. Similarly, other youth were participants of the RESOLVE by virtue of their placement in a residential detention facility during the duration of their sentence. In short, this challenge was about maintaining a rigorous data collection design despite substantial structural and demographic issues.

In confronting this challenge, the program evaluation team realized after careful consideration that there was no single “silver bullet” technique or process that could address this issue, so this challenge was addressed using a multidimensional layered approach to maintaining high data collection yields before, during, and after each evaluation administration.

Before each evaluation administration, the evaluation team worked in concert with program providers to ensure that youth were fully engaged in the program and had sufficient motivation to participate in the evaluation administrations. For example, care was taken to ensure that youth participants found the program relevant to their unique life situations as high-risk inner city youth. Likewise, program providers strived to ensure that youth understood the importance of the evaluation component of the program to increase the likelihood that they would (a) show up to the evaluation administration, and (b) exert a degree of motivation in completing the evaluation during the administration. Although empirical evidence demonstrates these initiatives were successful due to the relatively high yield, it is important to note that anecdotal evidence also suggests attempts to engage students in the program and the program’s evaluation were successful. At one location a student was so enthralled by the program that she

formed a “RESOLVE-type Club” at her high school, where the program had not been offered. During the actual evaluation administration a number of protocol were also introduced to maintain an acceptable data collection design. For example, proctors were empowered to verbally provide guidance to youth during administration sessions about the importance of the evaluations, with the intention being that if participants knew the value of the evaluations they would strive to provide full data on every question. Likewise, the evaluation administration protocol was expanded to include a step in which proctors verified that youth responded to every question on the evaluation. For those questions left blank by youth, proctors individually approached the youth to confirm that the youth intentionally left the question blank, and asked that the youth complete the missing question if the youth felt comfortable providing a response.

Once the formal evaluation window was complete, the evaluation team continued to work to ensure an acceptable yield so a sufficient amount of data could be available to make valid inferences about the program’s effectiveness. First, the evaluation team crafted a number of rosters that displayed which youth still had to complete a pretest or posttest, and data from these youth were included in the evaluation where appropriate. Second, the evaluation team implemented a number of missing data techniques to augment the data collection yield in instances in which there was evidence a youth’s response was left unanswered by accident. Unlike most traditional evaluations, extensive resources, especially time and money, was allocated specifically to ensure the evaluation could garner an acceptable data collection yield.

Sensitivity issues surrounding the high-risk population. A third challenge that confronted the evaluation team of the RESOLVE program was the myriad sensitivity issues surrounding the high-risk population of youth served by the program. Many youth from high-risk populations

have personal histories that include parental and/or personal abuse, use of drugs and alcohol, or cognitive challenges. Because the RESOLVE program targeted sensitive issues related to abstinence behaviors, it was imperative to ensure that youth trusted those individuals who were delivering the curriculum. Likewise, it was important that youth participants felt free to respond truthfully to the evaluation, which required an understanding by the youth that the evaluation team respected the youths' privacy and was committed to protecting that privacy.

Privacy, trust, and respect are not new issues to the evaluation of the RESOLVE program; in fact, other abstinence education programs have confronted similar issues (e.g., Trenholm et al., 2008; Brandon, 2010). Any program that addresses sensitive issues such as pregnancy and STD prevention requires extraordinary dignity, compassion, and seriousness, and RESOLVE is no different in this respect. However, the youth involved in the RESOLVE program represent a population of individuals who have substantially different—and more challenging—life situations than youth found in other similar programs. As noted previously, the high-risk youth who comprise the RESOLVE target population come from significantly adverse conditions. Some were in foster care or other group home settings, others were in residential facilities because they were required to leave foster care, and still others were incarcerated in detention centers. Few if any youth had parental figures who instilled healthy life choices. Because a large proportion of RESOLVE's youth participants came from backgrounds that did not foster proactive behaviors, RESOLVE program providers had to engage students in “unlearning” many of the beliefs and behaviors they had learned from their prior parental figures and other role models. Clearly this caused a degree of dissonance among youth as instructors attempted to help participants understand that, if they follow the paths of their role models, they were likely to end up seriously disadvantaged and possibly harmed. For example, many youth

participants are themselves children of teen pregnancies, and RESOLVE program providers worked diligently—yet delicately—to help youth understand the undesirable impact of becoming pregnant at an early age. Likewise, a large number of youth in the program were already sexually active, and many had already become pregnant or contracted an STD at least once, which further made an all-abstinence message challenging to contend with. In short, the abstinence message of the RESOLVE program was particularly hard for these youth participants to hear because it was so contrary to everything they had known to that point, and program providers and evaluators had to be cognizant of this fact.

This challenge was addressed through the use of trusted service providers as administrators of evaluation instruments, an extensive training program for adult service providers, and strict controls on the confidentiality of evaluation data. Importantly, while the selection and training of program providers and the maintenance of confidentiality is standard for sensitive programs, what set the RESOLVE program apart from traditional evaluations is the implementation of these techniques in conjunction with a set of guiding values. Chief among these values was the understanding that everyone involved with providing the program and evaluating the program's results had to believe in the program's message about youth needing to find their own agency to make good choices. A second guiding value was the understanding that the program's central tenant was that it offered youth an option so they would know that there are alternatives to life as it was lived by their parental figures and other role models. Third, the program endeavored to foster the value of mutual trust and respect among youth and program providers. Taken together, the intentional steps taken by RESOLVE staff to maintain trust and confidentiality enabled youth participants to feel safe and trusted and to participate in the program and respond truthfully to the program's evaluation.

Creation of materials that are developmentally appropriate. The fourth primary challenge encountered by the RESOLVE team was the provision of program content and evaluations that were both rigorous *and* developmentally appropriate. Standard evaluation techniques tend to call for providing program participants with curricula that has been administered and tested in small-scale laboratory-like settings, and rolled out for operational use only once the program has been sufficiently vetted. Likewise, evaluations tend to rely on instruments that have been investigated in low-stakes academic settings; the implementation of new or substantially altered psychometric instruments is normally not an acceptable practice in large operational program evaluations.

The RESOLVE program evaluation team realized at an early stage of the evaluation process that the program's youth participants did not have the capacity to understand and sufficiently respond to standard abstinence education and self-esteem-focused curricula and evaluations. Many youth from high-risk populations, including many of those who participated in the RESOLVE program, function at educational levels below grade-level expectations and consequently have difficulty reading and comprehending text as well as following directions. Furthermore, many of RESOLVE's youth participants tend to present with extensive persistence and attention issues that make it challenging for this population to take the same curricula and evaluations that are completed by more traditional youth populations.

To understand the important cognitive and behavioral (attention/persistence) issues that confront the RESOLVE program's youth it is helpful to consider the background of these program participants. As noted previously, most youth in the RESOLVE program have neither a traditional nor a stable home environment, and this unique challenge is understandably

associated with cognitive and behavioral challenges. Additionally, these youth are being told to internalize a message that is contrary to what they may have been taught by previous role models, and this dissonance is difficult for youth to accommodate.

This challenge was addressed by re-writing evaluation instruments and program curriculum materials such that they were appropriate for, and interpretable by, the program's youth. This was done by systematically minimizing the demands placed by the curriculum and evaluation on reading skills and attention span while maintaining the integrity of the constructs being taught and measured. Curricula and evaluations were rewritten to include language that is common to the youth participants. Also, the curriculum was crafted such that (a) all youth participated in small groups, (b) the curriculum was spread out over time when possible, (c) extensive discussion and other "active" elements were implemented in the curriculum, and (d) care was taken to include experiential activities in lieu of lecture when possible. The evaluation was also adjusted to ensure it was appropriate for these high-risk participants. For example, the evaluation measures were amended to remove psychometrically unnecessary items and various items from well-regarded instruments were rewritten to minimize demands on reading skills such as comprehension and sentence difficulty. The actual test administration protocol was amended so that service providers were instructed on how to administer evaluations to youth who have varying levels of academic experience. By revising the curriculum and evaluation protocol to ensure it is developmentally appropriate for the RESOLVE youth participants the evaluation team clearly made decisions that do not strictly follow common evaluation protocol; however, the revisions were made to ensure the curriculum and evaluation were accessible to these youth, thereby maintaining the validity of the inferences that can be made about the program's effectiveness.

Conclusion

The challenges and recommended solutions reported above are a selection of the issues facing evaluation practitioners as they seek to make inferences about program effectiveness from educational programs that target high-risk youth populations. High-risk populations differ from traditional populations in that the high-risk populations often require extensive revisions to common evaluation protocol to ensure that reliable and valid inferences are made from evaluation data. To be clear, this is a balancing act: Traditional evaluation techniques are widely accepted because they tend to yield strong inferences about the effectiveness of the programs they are used to evaluate; however, when applied to high-risk populations they often produce relatively *weak* inferences unless they are amended to accommodate the intricacies of this unique group of individuals. When are revisions to common protocol appropriate? How much revision is appropriate? Although these questions are addressed in this paper as they pertain to the RESOLVE program, one would expect these questions will continue to pervade the critical dialogue surrounding the evaluation of high-risk populations. Evaluating programs that involve high-risk populations is not easy, but it is important. Examining the challenges that high-risk populations present to evaluators can lead to the development of new evaluation protocols that are more appropriate to this unique group of individuals.

References

- Bickerman, L. & Rog, D.J. (Eds.). (1992). *Evaluating mental health services for children*. San Francisco: Jossey-Bass.
- Brandon, P. R., Smith, N. L., Trenholm, C., & Devaney, B. (2010). Evaluation exemplar: The critical importance of stakeholder relations in a national, experimental abstinence education evaluation. *American Journal of Evaluation*, doi:10.1177/1098214010382769
- Christie, C. A., & Fleischer, D. N. (2010). Insight into evaluation practice: A content analysis of designs and methods used in evaluation studies published in North American evaluation-focused journals. *American Journal of Evaluation*, 31(3), 326-346.
- Denscombe, M. (2002). *Ground rules for good research: A 10 point guide for social research*. Buckingham, UK: Open University Press.
- Hopson, R.K. (Ed.). (2000). *How and why language matters in evaluation*. San Francisco: Jossey-Bass.
- Jemmott, J. B. III, & Fry, D. (2002). The abstinence strategy for reducing sexual risk behavior. In A. O'Leary (Ed.), *Beyond condoms: alternative approaches to HIV prevention (pp. 109-137)*. New York: Kluwer Academic/Plenum.
- W.K. Kellogg Foundation (1998). *W.K. Kellogg Foundation Evaluation Handbook*. W.K. Kellogg Foundation: Battle Creek, MI.
- Kirkpatrick, D. (1959a). Techniques for evaluating training programs. *Journal of the American Society for Training and Development*, 13, 3-9.
- Kirkpatrick, D. (1959b). Techniques for evaluating training programs: Part 2--learning. *Journal of the American Society for Training and Development*, 13, 21-26.

- Kirkpatrick, D. (1960a). Techniques for evaluating training programs: Part 3--behavior. *Journal of the American Society for Training and Development*, 14, 13-18.
- Kirkpatrick, D. (1960b). Techniques for evaluating training programs: Part 4--results. *Journal of the American Society for Training and Development*, 14, 29-32.
- Madison, A.M. (Ed.). (1992). *Minority issues in program evaluation*. San Francisco: Jossey-Bass.
- Parker R.N., & Hudley, C. (Eds.). (2006). *Pitfalls and pratfalls: Null and negative findings in evaluation interventions*. San Francisco: Jossey-Bass.
- Perloff, R. & Perloff, E. (Eds.). (1980). *Values, ethics, and standards in evaluation*. San Francisco: Jossey-Bass.
- Rog, D.J. (Ed.). (1991). *Evaluating programs for the homeless*. San Francisco: Jossey-Bass.
- Rossi, P.H., Lipsey, M.W., & Freeman, H.E. (2004). *Evaluation: A Systematic Approach, Seventh Edition*. Thousand Oaks, CA: Sage Publications.
- Sanders, J. (1994). *The program evaluation standards: How to assess evaluations of educational programs*. Thousand Oaks, CA: Sage Publications.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Smith, N.L. (Ed.). (1981). *Federal efforts to develop new evaluation methods*. San Francisco: Jossey-Bass.
- Thompson-Robinson, M., Hopson, R., & SenGupta, S. (Eds.). (2004). *In search of cultural competence in evaluation: Toward principles and practices*. San Francisco: Jossey-Bass.
- Trochim, W. M. (2006). *The research methods knowledge base, 2nd edition*. Retrieved April 20, 2009, from <http://www.socialresearchmethods.net/kb/>

Trenholm, C., Devaney, B., Fortson, K., Clark, M., Quay, L., & Wheeler, J. (2008). Impacts of abstinence education on teen sexual activity, risk of pregnancy, and risk of sexually transmitted diseases. *Journal of Policy Analysis and Management*, 27(2), 255-276.

Acknowledgements

The authors wish to thank the many people who contributed to the success of the evaluation of the RESOLVE program: The executive leadership of the Jewish Child Care Agency, in particular, Richard Altman, CEO; Elizabeth Schnur, Senior Vice President of Quality Management and Planning; Amy Cohen, Director of Grants Management; Alexandra Bowie, Director of Quality Improvement and Policy; Rimi Nagi, Grants Management Assistant; Richard Hucke, Division Director of Coordinated Community Services; RESOLVE program Coordinators Charletta Robinson and Cadine Soman, along with their Administrative Assistants Cynthia Puello and Yolanda Smoak and group facilitators; members of the Community Advisory Board; William Jere Hough, Jr.; and the youth participants along with their parents and foster parents. This research was supported by Grant #90AE0225 from the U.S. Department of Health and Human Services Administration for Children & Families, Family and Youth Services Bureau.