

**Collaborative evaluation of a high school prevention curriculum: How methods of collaborative evaluation enhanced a randomized control trial to inform program improvement.**

By: Muhsin Michael Orsini, David L. Wyrick and Jeffrey J. Milroy

[Orsini, M. M.](#), [Wyrick, D. L.](#), & [Milroy, J. J.](#) (2012). Collaborative evaluation of a high school prevention curriculum: How methods of collaborative evaluation enhanced a randomized control trial to inform program improvement. *Evaluation and Program Planning*, 35(4), 529-534.

**Made available courtesy of Elsevier:**

<http://www.sciencedirect.com/science/article/pii/S0149718911001212>

**\*\*\*Reprinted with permission. No further reproduction is authorized without written permission from Elsevier. This version of the document is not the version of record. Figures and/or pictures may be missing from this format of the document. \*\*\***

**Abstract:**

Blending high-quality and rigorous research with pure evaluation practice can often be best accomplished through thoughtful collaboration. The evaluation of a high school drug prevention program (All Stars Senior) is an example of how perceived competing purposes and methodologies can coexist to investigate formative and summative outcome variables that can be used for program improvement. Throughout this project there were many examples of client learning from evaluator and evaluator learning from client. This article presents convincing evidence that collaborative evaluation can improve the design, implementation, and findings of the randomized control trial. Throughout this paper, we discuss many examples of good science, good evaluation, and other practical benefits of practicing collaborative evaluation. Ultimately, the authors created the term pre-formative evaluation to describe the period prior to data collection and before program implementation, when collaborative evaluation can inform program improvement.

**Keywords:** collaboration | collaborative planning | collaborative evaluation | drug prevention | randomized control trial | evaluation practice | pre-formative evaluation

**Article:**

1. Introduction

Often there is confusion when attempting to distinguish between high-quality practices of research and evaluation. Within the social and applied sciences, most research is considered basic in that it focuses on generalizing findings to promote understanding of a phenomenon or theoretical construct. For example, educational psychologists study student motivation among adolescents and sociologists examine the influence of family economic status on career choice.

On the other hand, evaluation focuses on investigating the effectiveness of a particular program in a particular place at a particular time. Although it may be possible to identify causal linkages, generalizing findings to other contexts is not usually the primary focus in the evaluation. In general, the evaluator investigates program implementation and outcomes to learn about program processes and collect information that be used for program improvement. Rarely does an evaluator generalize findings to learn about theoretical constructs of interest to other researchers.

The purpose of this article is to showcase a collaborative evaluation project that utilized high-quality and rigorous research and evaluation practice, in essence, blending a prescriptive research design with methods of collaborative evaluation (Orsini, Wyrick, O'Sullivan, & Hansen, 2005). The project began as collaboration between the program developers at Tanglewood Research (Tanglewood) and the external evaluators at Evaluation, Assessment, & Policy Connections (EvAP) at the University of North Carolina at Chapel Hill, School of Education. Tanglewood is dedicated to developing, testing, training, and marketing highly effective educational materials for preventing drug use, violence, delinquency, and premature sexual activity among teens. EvAP conducts evaluations, trains evaluators, and aspires to build the evaluation capacity and effectiveness of private, public and nonprofit organizations using the methods of collaborative evaluation (O'Sullivan, 2004).

Tanglewood has very specific goals when developing new drug prevention programs. Legislation adopted by funding agencies such as the Federal Safe and Drug Free Schools Program suggested that schools adopt evidence-based programs that have been evaluated and deemed to be effective. When assessing the effectiveness of their evidenced-based programs, Tanglewood often hires an independent evaluator and implements a randomized controlled trial (RCT). Additionally, the Substance Abuse and Mental Health Services Administration (SAMHSA) has established the National Registry of Evidence-based Programs and Practices (NREPP) to connect members of the public to intervention developers. The review of interventions includes an assessment of the intervention's evaluation outcomes, quality of the research/evaluation, targeted population, readiness for dissemination, and replication studies. SAMHSA places a strong emphasis on the quality of the research by assessing the reliability of measures, validity of measures, intervention fidelity, missing data and attrition, potential confounding variables, and appropriateness of analysis. Tanglewood operates with the understanding that the likelihood of achieving a high quality program evaluation across the areas listed above is best accomplished through the use of an independent evaluator and rigorous methods such as the randomized control trial.

Tanglewood was awarded a grant from the National Institute on Drug Abuse to develop and evaluate a high school drug prevention program – All Stars Senior, to be implemented in North

Carolina, Nebraska, and Missouri public schools. The curriculum consists of activities organized around the 10 traditional content areas of school health education (Centers for Disease Control and Prevention, 2010). Program materials were intended to provide high school health teachers with supplemental activities to be delivered in health class. Instructional activities were designed to be interactive and engaging, using small group activities, games, discussions, and worksheets that asked students to apply knowledge to real-world tasks. For each content area, between 15 and 30 activities are available. Each activity targets changing a mediating variable previously identified as a predictor of either substance use onset (McNeal & Hansen, 1999) or early self-initiated substance use cessation (Hansen & McNeal, 2001). Targeted mediating variables include: normative beliefs, commitment to avoid risky behavior or participate in healthy behavior, perceptions that risky behaviors would interfere with desired lifestyles and that healthy behaviors would contribute to desired lifestyles, the development of goal setting, decision making, stress management, and peer pressure resistance skills.

EvAP was awarded the contract to serve as the independent evaluator for the All Stars Senior program. During initial planning meetings, EvAP communicated to Tanglewood information from the Guiding Principles for Evaluators of the American Evaluation Association (AEA, 1995). EvAP would adhere to the highest appropriate technical standards in conducting the evaluation and explore with the client both the shortcomings and strengths of various approaches that might be employed to answer evaluation questions. The spirit of collaboration came to define a reciprocal relationship and was set in motion at the very start of this project. In this article both client and evaluator will share lessons learned while working toward the goal of blending collaborative evaluation and the RCT.

## 2. Recruitment

Typically, when planning for school-based projects, Tanglewood would meet with the principal and teachers at potential schools after soliciting approval and support from school district administrators. During this project, the collaboration between EvAP and Tanglewood led to the development and implementation of a systematic recruitment process. EvAP recommended that recruitment efforts would be more effective if a team comprised of both program and evaluation specialists visited each high school to make a presentation regarding the specifics of the project (program implementation and evaluation). EvAP prepared a detailed recruitment protocol and information packet for each site. The protocol outlined a systematic method for recruiting high schools. First, the district level administrator in charge of high school health education was identified and contacted by telephone. A specific script was followed for these conversations to briefly introduce the program and evaluation team, the broad aims of the project, and that funding was provided by the National Institute on Drug Abuse (NIDA). Permission was then

requested to send the administrator a packet that included more detailed information about the All Stars Senior program; procedures for implementation of the program and evaluation activities; expectations for participating schools, teachers, and students; and incentives for participation. At this time, permission to contact principals at each of the high schools in the district was requested.

It is important to note that three school districts required the program and evaluation team to prepare and submit proposals to be reviewed by district level research committees. In the end, each of these proposals was approved. Once permission was granted by the district administrator, an email was sent to each principal informing them that permission was received to contact them and that district administrators had identified this project as an opportunity in which they may be interested. Several days/times were recommended for a brief conference call that followed the same script that was used with the district administrator. At the conclusion of each conference call a face-to-face site visit was scheduled to meet with the principal, health teachers, and any other appropriate faculty/staff who the principal thought should be included.

Prior to each site visit, the program and evaluation team conducted some basic fact finding about the school and community. For example, the success of athletic teams and any district, state, or national academic awards that had been earned by the school were identified. Informal conversation at the beginning of each meeting established familiarity and helped school personnel understand that their participation was essential to the success of this project.

During site visits a detailed information packet was delivered to school personnel in attendance and reviewed. The packet included several copies of an All Stars Senior brochure describing the curriculum, parental consent form, student survey, and project timeline describing evaluation activities. After discussion and responding to questions, the principal was asked to endorse a memorandum of understanding that succinctly stated the roles and responsibilities of the school, health teachers, students, and program and evaluation teams. In the end, the program and evaluation team, school leaders, and health teachers understood that they had made commitments to each other for which they would be held accountable. For example, financial incentives that were promised were not delivered to schools by Tanglewood until after student survey data were received by EvAP. This was a very important factor in the success of recruitment and data collection. Whereas, two high schools that had committed to the project withdrew prior to data collection due to changes in school administration, 24 schools and 50 teachers agreed to participate in this project; remarkably, 100% of them fulfilled their yearlong commitment. Compared to previous school-based projects implemented by Tanglewood, this was an extremely

large number of schools and teachers to recruit. Anecdotal feedback from principals and teachers indicated that the success in recruiting was directly related to the site visits.

In the end, the client experienced and reported high levels of cooperation from participants. For example, there was easy access to teachers and classrooms, thorough understanding of project purposes by participants, and minimal attrition of schools and teachers.

### 3. Evaluation design

Tanglewood conducted a small-scale evaluation of All Stars Senior during Phase I of the NIDA grant. The design and all instruments for the Phase I evaluation were developed internally. During this project, curriculum specialists from Tanglewood and evaluation specialists from EvAP worked together to negotiate a study that remained true to the purpose of program evaluation and incorporated methods of the randomized control trial. Although the purpose of the Phase I evaluation did not require an RCT, the client reported that more reliable and valid instruments were implemented during this project. For example, variables were derived from a more advanced understanding of program theory by the evaluator, additional moderating variables were identified by the evaluator for measurement, and instruments were more participant-friendly.

One of the most important benefits from the collaborative relationship between EvAP and Tanglewood was experienced during design of the evaluation and development of data collection instruments and protocols. This result is consistent with previous research of participatory evaluation processes (Cousins, Donohue, & Bloom, 1996; and Cousins & Whitmore, 1998). Furthermore, comments from the National Institutes of Health study section that reviewed the original grant proposal (Center for Scientific Review, 2000) highlighted strengths of the evaluation design. First, reviewers reported that “the research is methodologically and statistically well designed.” Second, reviewers noted a “focus on the fidelity of program implementation” and valued that “fidelity will be treated as a variable and multiple measures will be used to measure program fidelity.” Lastly, reviewers stated that “most importantly, the investigators have chosen to engage an external evaluation team to independently evaluate the success of the program, which will ensure that bias does not affect program measurement.”

A critical first step in the collaborative process is to mutually decide on the purpose of the evaluation. This often seems obvious to program personnel as it did to Tanglewood when they first approached EvAP about leading this evaluation. Tanglewood expressed that they wanted to

use rigorous methods of RCT to determine if All Stars Senior was successful in achieving its intermediate and long-term outcomes. Tanglewood had carefully adhered to a specific theoretical model when developing All Stars Senior. As illustrated in Fig. 1, the program was carefully designed to directly change a number of characteristics that reduce risk and increase protection from participating in high-risk behaviors.



Fig. 1.

Theoretical model of All Stars Senior.

After some initial discussions and strategic questions from EvAP, it became apparent to Tanglewood that if the only objective of the evaluation was to assess components of the program model, that there would be some missed opportunities. Therefore, the purpose of the evaluation was expanded to include an investigation of factors that influence program implementation as well as outcomes (student attitudes and behaviors). The Evaluation Crosswalk (O'Sullivan, 1991) in Fig. 2 presents the evaluation questions investigated and data collection strategies implemented to achieve this newly conceived purpose.

	Survey of Students x 3	Focus Group w/Students	Focus Group w/Teachers	Observation of Instruction	Teaching Logs	Case Study
<b>Program Outcomes</b>						
1. To what extent does program impact attitudes, substance use, and other high-risk behaviors?	X	X				
2. To what extent does program impact various mediating variables?	X	X				
3. To what extent do various moderating variables impact outcomes?	X	X	X	X	X	
<b>Program Implementation</b>						
1. What program factors support success?			X			X
2. What environmental factors support success?			X			X

Fig. 2.

Evaluation crosswalk.

As can be seen in Fig. 2, factors related to program implementation and outcomes were assessed with a minimum of two different data collection strategies. EvAP proposed a mixed-methods design to increase the validity of measurement and provide information to assist in the interpretation of findings. Table 1 and Table 2 presents the variables of interest related to program implementation and outcomes derived from evaluation questions that were assessed in this project.

Table 1.

Program implementation variables assessed in the project.

Program factors	Environmental factors	Moderating variables
Teacher training	School culture	Fidelity
Teaching materials	Teacher certification	Dosage
Student materials	Other school-wide initiatives	Teacher–student rapport

Table 2.

Program outcome variables assessed in the project.

Mediating variables	Outcome variables
Commitment	Substance use
Normative beliefs	Risky behavior
Decision making skills	Problem school behavior
Lifestyle incongruence	
Resistance skills	
Beliefs about consequences	
Goal setting skills	
Stress management skills	
School attachment	

EvAP had conducted many school-based projects prior to collaborating with Tanglewood on this project. Their experience taught them that high schools are much more agreeable to research projects that minimize burdens for the school, students, and parents. One task that can be particularly burdensome involves gathering active informed consent from parents or guardians for every student. Therefore, EvAP advised that it would be best to apply for a waiver of active informed consent from the Institutional Review Board. EvAP relied on their extensive experience conducting school-based projects to prepare the application. In the end, anecdotal evidence reported by school principals suggested that using implicit consent procedures led to a larger number of students than usual whom were eligible to participate in evaluation activities, as compared to their previous experiences.

During recruitment visits, school principals expressed concerns about the use of student surveys written in English because there were a significant number of Spanish-speaking students at their schools. Therefore, the program and evaluation team deemed it necessary to translate consent forms and the student surveys, and present both English and Spanish versions of the survey to students who were pre-identified by principals as Spanish-speaking. EvAP had previous experience translating data collection instruments and protocols from English to Spanish so they



led this effort. It can be assumed that this resulted in less missing data because students who were learning English as a second language were able to respond to the survey items in their preferred language.

Not spending sufficient time developing and pilot testing data collection instruments can result in the collection of data that does not meet standards of validity (Teijlingen & Hundley, 2001). Therefore, the program implementation and evaluation team devoted considerable effort to pilot testing the student survey used during previous evaluations of All Stars curricula. This provided evidence to determine which items were appropriate and which variables of interest needed to be measured by more reliable and valid items. For example:

- Previously Tanglewood had created a series of survey items to measure prosocial bonding that did not demonstrate high reliability or load on the same factor during the pilot test. In lieu of these findings, EvAP clarified exactly what Tanglewood wanted to measure. After discussion it was discovered that Tanglewood was actually interested in measuring what is referred to in the educational literature as school attachment. Thus, EvAP identified a school attachment scale that had been previously validated and widely used.
- Previously Tanglewood had used a single scale to measure decision making skills. However, factor analysis of pilot test data suggested that a two factor solution was a better fit. In the end, some items were eliminated and the remaining items were separated and used to measure two variables, decision making skills and impulsive decision making.

To achieve the best response rate as possible, EvAP and Tanglewood collaborated to modify the student survey to be more participant-friendly. For example, the format of the survey was improved by reorganizing substance use items. An answer choice grid was used instead of presenting separate questions for each substance, reducing the amount of time necessary for completing the survey. This was intended to decrease the burden of data collection for participants. Additionally, EvAP consulted the literature to identify a more comprehensive list of substances and expanded the number of associated nicknames presented to participants. This was intended to improve understanding of survey items among the adolescent participants.

A common challenge for research and evaluation projects is the successful collection and acquisition of data. It is not unusual for data to be successfully collected but somehow lost in the acquisition process (i.e., getting the data to the evaluator). During collaborative planning the team anticipated this challenge and assigned a specific person from EvAP to follow-up with teachers about retrieving data. One particular school reported that they “lost” all student survey data for one of the posttests. Instead of accepting the idea that these surveys were lost and had somehow been discarded, EvAP made phone calls to several individuals at the school trying to

track down the surveys. Eventually, the boxes of surveys were located in a janitor's closet and returned for data entry and analysis. While there was certainly student-level attrition, 100% of requested data were collected and returned.

#### 4. Program implementation

When conducting school-based projects, Tanglewood would meet with teachers at schools to carry out training. During this project, a curriculum specialist from Tanglewood delivered curriculum training and an evaluation specialist from EvAP delivered an evaluation briefing during regional meetings. As discussed below, collaboration in the design and implementation of teacher trainings contributed to improved program fidelity and compliance than experienced in Phase I. For example, the Phase I study Tanglewood observed that high school health teachers often modify lesson plans. However, in this project, Tanglewood reported that modification of lessons by teachers was consistent with program objectives more often than usual and a greater number of lessons were delivered to students.

In preparing for this project, the evaluation team performed a thorough review of the All Stars Senior curriculum. The curriculum was originally organized simply by content (e.g., alcohol, tobacco, and other drugs, personal health, and injury prevention). The evaluation team recommended that the curriculum be further organized within each content area by mediating variable (e.g., normative beliefs, commitment, and decision making skills). This was intended to provide teachers a constant reminder of the mediating variable targeted by the activity so that they would be encouraged to adhere more closely to the lesson plan, thereby increasing program fidelity.

In a project of this magnitude involving a comprehensive curriculum such as All Stars Senior, quality teacher training is essential. The evaluation team spent a considerable amount of time discussing specific content for the training, the appropriate length of the training, who should conduct the training, and the most appropriate methods for content delivery. In the end, it was mutually decided that trainings should be conducted by representatives of both Tanglewood and EvAP so that health teachers could distinguish between the two organizations responsible for program implementation and evaluation. Additionally, EvAP strongly recommended that feedback be solicited from health teachers at the end of training to identify unforeseen concerns and problems, and inform improvement of training.

Two day trainings were scheduled in North Carolina, Nebraska, and Missouri. When finalizing the schedule for the training it was decided that 1½ days would be dedicated to introducing the All Stars Senior curriculum and a ½ day would be dedicated to introducing evaluation activities. Reserving 25% of training time for discussion of evaluation may seem excessive; however, the evaluation team realized that the success of all data collection was dependent upon the compliance of health teachers to data collection protocols. Teachers were informed about the purpose of the project, their roles and responsibilities, specific details and rationale for evaluation activities, and who to contact for support. There was concern that teachers would be confused by instructions to contact Tanglewood with questions regarding the All Stars Senior program and EvAP with questions regarding evaluation activities, but in the end teachers reported during interviews that they appreciated being able to discuss their topic of concern with the person(s) responsible for that task. It is important to note that occasionally Tanglewood received questions about evaluation activities and EvAP received questions about program implementation. Due to the collaborative nature of this project, both program implementation and evaluation teams were equipped to answer questions immediately without the need for consultation. The collaboration improved the ability of both organizations to provide efficient customer service.

All Stars Senior provides health teachers with a large selection of activities, more than they could possibly implement during health class. To encourage implementation of the curriculum in classrooms, the evaluation team recommended that health teachers be required to deliver at least 2–3 All Stars Senior activities per week of instruction. EvAP also recommended that Tanglewood identify a few activities that were essential for preventing substance use and other problem behaviors among high school students and require that health teachers deliver these essential activities. In the end, these requirements were formalized in the memorandum of understanding and were emphasized in both training materials and curriculum guide. Establishing minimum requirements for implementation of All Stars Senior in the classroom was intended to increase the likelihood of a measurable effect.

## 5. Lessons learned

The RCT is generally accepted as the most reliable form of scientific evidence. It is the gold standard for demonstrating the efficacy and/or effectiveness of a program. Efficacy trials are explanatory in that they test the program in highly controlled evaluation settings while effectiveness trials are pragmatic in that they test the program in everyday conditions. Regardless of whether the reason for designing a program evaluation around an RCT is explanatory or pragmatic, the use of the RCT requires rigorous methods such as multiple comparison groups and random assignment. This is often the source of a common conundrum. The multiple goals

and methods of the RCT are often viewed in direct contrast with the primary purpose of program evaluation which is to inform program improvement. Considering the fact that success of a program evaluation is directly related to the utility of the evaluation results, the contrast seemingly becomes even more evident.

Collaborative evaluations are recognized for maximizing the utilization of evaluation results. So the primary lesson of this study is that the collaborative process between client and evaluator can result in an evaluation design that can effectively blend the requirements of an RCT to demonstrate program effectiveness with more traditional evaluation methods to inform program improvement.

In this study, the program developers (Tanglewood Research) came to the evaluators (EvAP) with the seemingly contradictory needs described above: (a) the need for an independent evaluation to demonstrate program effectiveness, and (b) the need to pilot new program materials and gather data to inform and guide program improvement. It is important to acknowledge the critical first step taken by Tanglewood Research. That is, to engage an evaluator who practices collaborative evaluation and could design an evaluation capable of meeting both of these important needs. Two lessons are evidenced here. First, evaluators need to do a good job communicating their philosophical approach to evaluation when marketing their services so that clients are aware of the expertise that is available to assist with program evaluation needs. Secondly, a simple but important lesson is that the client was afforded the opportunity to clearly articulate these needs and to work collaboratively with the evaluator to design an appropriate evaluation.

There are many possible outcomes of any collaboration. The last lesson is that when collaborations are attempted between parties who are poorly matched or are not willing to accept alternative viewpoints, poor outcomes are almost guaranteed. The productive atmosphere of the All Stars Senior collaborative planning sessions and enduring relationships that emerged, are evidence of the positive outcomes that can result from collaborative evaluation done right. Throughout this paper, we have discussed many examples of good science, good evaluation, and other practical benefits of practicing collaborative evaluation.

Ultimately, the authors created the term pre-formative evaluation to describe outcomes of this collaboration and reciprocal relationship. Prior to data collection and before program implementation, collaborative evaluation was actively informing program improvement. It is

during this preformative stage that evaluator and client can make the critical decisions of whether the evaluator/client match is a good one and whether or not to move forward collaboratively to conduct the program evaluation.

### Acknowledgements

We want to thank the rest of the program implementation and evaluation teams who collaborated so well together throughout this project: Bill Hansen, Rita O'Sullivan, Allan Steckler, and Denise Halfors. We would also like to express appreciation to participating high schools, teachers, and students whose cooperation made this project possible.

This project was supported in part with funding from the National Institute on Drug Use; grant number R44 DA12302, William B. Hansen, Principal Investigator.

### References

AEA, 1995. American Evaluation Association. Guiding principles for evaluators. *New Directions for Program Evaluation*, 34 (1995), pp. 19–26

Centers for Disease Control and Prevention, Division of Adolescent and School Health. (2010). School health education resources. Retrieved from: <http://www.cdc.gov/healthyyouth/> Accessed 03.08.10.

Center for Scientific Review, National Institutes of Health (2000). Summary statement [Review of the grant proposal Drug prevention in high schools: All Stars, Sr.]. Washington, DC: Center for Scientific Review, National Institutes of Health.

J. Cousins, J. Donohue, G. Bloom. Collaborative evaluation in North America: Evaluators' self-reported opinions, practices, and consequences. *Evaluation Practice*, 17 (3) (1996), pp. 207–226

J. Cousins, E. Whitmore. Framing participatory evaluation, in: E. Whitmore (Ed.), *Understanding and practicing participatory evaluation*, New directions for evaluation, Vol. 80, Jossey-Bass, San Francisco (1998), pp. 5–23

W.B. Hansen, R.B. McNeal. Self-initiated cessation from substance use: A longitudinal study of the relationship between postulated mediators and quitting. *Journal of Drug Issues*, 31 (2001), pp. 957–974

R.B. McNeal, W.B. Hansen. Developmental patterns associated with the onset of drug use: Changes in postulated mediators during adolescence. *Journal of Drug Issues*, 29 (2) (1999), pp. 381–400

M.M. Orsini, D. Wyrick, R.G. O'Sullivan, W. Hansen. Blending collaborative evaluation and the randomized controlled trial: Serving the diverse needs of client and evaluator. Paper presented at the annual meeting of the American Evaluation Association, Toronto, Ontario, Canada, October (2005)

R.G. O'Sullivan. *Practicing evaluation: A collaborative approach*. Sage Publications, Thousand Oaks, CA (2004)

R.G. O'Sullivan. Improving evaluation design and use through the “evaluation crosswalk” method. *National Forum of Applied Education Research Journal*, 4 (1991), pp. 43–49

E.R. Teijlingen, V. Hundley. The importance of pilot studies. *Social Research Update*, 35 (2001), pp. 1–4