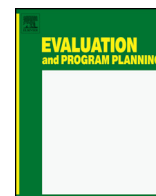




ELSEVIER

Contents lists available at ScienceDirect

Evaluation and Program Planning

journal homepage: www.elsevier.com/locate/evalprogplan

Evidence-based programs registry: Blueprints for Healthy Youth Development

Sharon F. Mihalic^{a,*}, Delbert S. Elliott^{b,1}^a Center for the Study and Prevention of Violence, University of Colorado Boulder, 483 UCB, Boulder, CO 80309, United States^b Institute of Behavioral Science, University of Colorado Boulder, 483 UCB, Boulder, CO 80309, United States

ARTICLE INFO

Article history:

Available online 19 August 2014

ABSTRACT

There is a growing demand for evidence-based programs to promote healthy youth development, but this growth has been accompanied by confusion related to varying definitions of evidence-based and mixed messages regarding which programs can claim this designation. The registries that identify evidence-based programs, while intended to help users sift through the findings and claims regarding programs, has oftentimes led to more confusion with their differing standards and program ratings. The advantages of using evidence-based programs and the importance of adopting a high standard of evidence, especially when taking programs to scale, are described. One evidence-based registry is highlighted—Blueprints for Healthy Youth Development hosted at the University of Colorado Boulder. Unlike any previous initiative of its kind, Blueprints established unmatched standards for identifying evidence-based programs and has acted in a way similar to the FDA – evaluating evidence, data and research to determine which programs meet their high standard of proven efficacy.

Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

Prior to 1990, the general consensus in the research community about the effectiveness of prevention programs was that “nothing worked,” or to be more precise, nothing had been demonstrated in evaluations of programs and practices to be effective in preventing delinquency, antisocial behavior or dysfunctional, health compromising behavior (Martinson, 1974; Romig, 1999; Sechrest, White & Brown, 1979). However, over the last two decades, there have been major advances in both evaluation research and program design and development. This work has provided a rich body of evidence demonstrating that some programs and practices are effective, both for preventing the onset of problem behaviors and for successfully intervening with those caught up in these types of behavior (Greenwood, 2006; Institute of Medicine, 2008; Sherman, Farrington, Welsh, & McKenzie, 2002). Moreover, these programs often have positive effects on other important outcomes such as mental health, academic achievement, parenting practices and family wellbeing, and employment. This change in findings

about the effectiveness of prevention programs and practices is the result of both major improvements in the quality of evaluation research and improved program design and implementation.

We now have a better understanding of what does and does not work, and this has led to a new interest in identifying and implementing programs that have been demonstrated by rigorous evaluations to be effective. This current drive for proven, evidence-based programs has also been fueled by huge financial deficits at both the federal and state levels, leading to serious consideration of the high costs of violence, crime, drug abuse, school dropout and other problem behaviors and the efficiencies associated with investments in more cost-effective, proven programs and practices. In 2002, the White House encouraged all federal agencies to support evidence-based programs and to discontinue programs without evidence of effectiveness (Office of Management Budget, 2001; 2002), and it is now common practice that federal and state funding for prevention programs be restricted to evidence-based programs and practices.

This paper seeks to better inform policymakers, practitioners and citizens about the importance and advantages of using evidence-based programs to improve the life course of children, taking a closer look at the Blueprints for Healthy Youth Development registry as one source of important information on this topic.

* Corresponding author. Tel.: +1 303 492 2137.

E-mail addresses: Sharon.mihalic@colorado.edu (S.F. Mihalic),Delbert.elliott@colorado.edu (D.S. Elliott).¹ Tel.: +1 303 735 1065.

1.1. Defining evidence-based programs

An evidence-based program is a set of coordinated services/activities that demonstrate effectiveness [on some desired outcome] based on research (Children's Services Council, n.d.). Most researchers agree that the evidence of program effectiveness should minimally come from quasi-experimental or experimental evaluation. Randomized experiments are the "gold standard" for determining the effectiveness of a program (Campbell & Boruch, 1975; Shadish, Cook, & Leviton, 1991). Some argue that a higher standard should be placed on programs that will be taken to scale, such as requiring randomized controlled trials (Coalition for Evidence-Based Policy, 2014; Elliott, 2013) or evidence of sustained effects and replication (Elliott & Mihalic, 2004).

A number of agencies and groups have developed standards for assessing the research for the effectiveness of programs in order to designate them as evidence-based. However, the standards adopted by each agency differ, with some applying a more rigorous standard than others. For example, some agencies that rate programs will only accept a randomized controlled trial as sufficient evidence (<http://evidencebasedprograms.org>), while some will accept both randomized controlled trials and closely matched quasi-experimental designs (<http://blueprintsprograms.com>). The higher standards, such as randomization, replication, and sustainability, will result in fewer programs, but it is critical that there be a high degree of confidence in the effectiveness of a program before endorsing and taking a program to scale.

The problem is that a lower standard comes with a greater risk of failure when programs are subsequently implemented on a wide scale. For example, evaluations conducted with RCTs have, in a number of instances, invalidated earlier findings from studies with quasi-experimental comparison group designs. Examples include hormone replacement therapy which was once a recommended treatment for postmenopausal women, based upon comparison group studies, until two large-scale randomized controlled studies showed that it increased the risk of coronary heart disease, stroke, and breast cancer; dietary fiber to prevent colon cancer was shown to have no effect; and an oxygen-rich environment for premature infants was shown to increase blindness (Baron, 2007).

A number of "design replication" studies have been carried out to examine whether and under what circumstances comparison-group studies can replicate the results of randomized controlled trials. These studies test comparison-group methods against randomized methods by first comparing the outcomes of the program group to a randomly assigned control group, and next comparing the same program participants with a comparison group selected through methods other than randomization. Twelve of these studies have been summarized by Cook, Shadish, and Wong (2008). Their review suggests that comparison group studies without close matching often produce inaccurate estimates of an intervention's effects. This is true even when statistical techniques are used to adjust for observed differences between the two groups. Often studies match only on demographic variables, and these studies consistently fail to reproduce the results of experiments. Comparison group designs are more likely to produce valid results when there is careful matching of the treatment and comparison groups at pretest, especially on the pretest measures of the outcome and geographic location.

The evidence used to inform policy decisions must be scientifically valid. Randomized controlled trials are first and foremost in generating this evidence, followed closely by matched comparison designs. Non-equivalent comparison group designs or methods that fail to use a control group do not provide an acceptable standard of evidence, as they often produce

erroneous results. Other factors in design and implementation of an evaluation must also be considered to ensure that the evaluation is producing valid results. These include, but are not limited to: adequate sample size, baseline equivalence, low attrition, lack of differential attrition, valid outcome measures, appropriate unit of analysis, intent to treat analysis, and appropriate statistical techniques. The Society for Prevention Research has adopted a similar set of standards that must be met if a program or policy is to be called tested and effective (Society for Prevention Research, n.d.).

The quality of evidence is not the only consideration in defining an evidence-based program. If these programs are to be replicated, there must be specificity in the program description that clearly shows how its theoretically grounded components produce the intended impact. It is, therefore, important to identify the outcomes the program is designed to change and the specific risk, protective, and promotive factors that will mediate that change. It is also important to designate the targeted population, which should not be based upon assumption, but upon evidence of the program's success with that population. Theoretically driven programs also involve detailed instructions on how to deliver the intervention, duration of the intervention, and amount of training required. Failure to implement the program within the specified guidelines often results in smaller or null effects (Mihalic, 2004). While some may question the importance of using theoretically driven programs, recent studies indicate that interventions which make extensive use of theory tend to have larger effects on behavior than interventions that make less or no use of theory (Taylor, Conner, & Lawton, 2012; Webb, Joseph, Yardley, & Michie, 2010).

Widespread dissemination of programs with evidence of effectiveness from poorly designed studies, as well as implementation of programs with poor fidelity, is a waste of limited funds and undermines the public confidence in prevention science when the outcomes that were promised are not achieved.

2. Why has policy changed over the last decade to support evidence-based programs?

Budget shortfalls at national and local levels have created a need for greater efficiency and accountability in systems working with children and youth. Despite tremendous outlays of money each year for support services to families and youth, research is not being used with sufficient frequency, intensity and quality to impact human services and has not provided the full potential benefits to consumers and communities (Baron, 2012; Sawhill & Baron, 2010). Baron (2012) uses as an example Department of Education data showing that "reading and math achievement of 17-year-olds—the end product of our K-12 educational system—has not improved over 40 years, despite a 90 percent rise in public spending per student (adjusted for inflation)." In the same report, he also states that "in education, although the college graduation rate has risen, the high school graduation rate peaked around 81 percent in the early 1970s. Since then, it has been stuck between 75 and 80 percent."

However, there are examples showing that when evidence-based programs are integrated into these systems, taxpayers enjoy cost savings and youth benefit from better outcomes. In 2004, the Florida Legislature voted to initiate the Department of Juvenile Justice's Redirection project to address the growing number of juvenile offenders who were being committed to residential facilities for non-legal violations of probation. The Redirection project diverted, or redirected, these youth from residential placement to evidence-based, community-based treatments, relying on three programs (Functional Family Therapy, Multi-systemic Therapy, and Brief Strategic Family Therapy). During the

eight-year period through 2012, Redirection saved the Juvenile Justice System \$183 million. Additionally, Redirection youth, compared to matched youth in residential commitment with the same risk profiles, had 19 percent fewer subsequent adjudications and 31 percent fewer subsequent placements. These reductions came with an average cost savings of \$22,000 per youth served in Redirection. The program has been consistently shown to enhance public safety and save taxpayer dollars (Redirection, 2012).

Pennsylvania also invested in evidence-based programs, especially Blueprints programs, with costs and benefits assessed for seven of those programs (Big Brothers Big Sisters, LifeSkills Training, Multidimensional Treatment Foster Care, Multisystemic Therapy, Functional Family Therapy, Nurse Family Partnership, and Strengthening Families). Using a conservative and widely accepted methodology developed by the Washington State Institute for Public Policy (WSIPP, 2013) and applied to data on the effectiveness of the programs in Pennsylvania, these programs achieved a return on investment of \$317 million in reduced corrections costs, welfare and social services burden, drug and mental health treatment, and increased employment and tax revenue. That figure translates into a return between \$1 and \$25 per dollar invested (Jones, Bumbarger, Greenberg, Greenwood, & Kyler, 2008).

After decades of reacting to problems with short-term, ineffective, or unevaluated approaches, and in some instances even believing in the pessimistic perspective that nothing works (Martinson, 1974; Regnery, 1985; Sechrest et al., 1979), we as a society have become more proactive in addressing the behavioral problems of youth and now pursue programs that have proven to demonstrate positive outcomes. Successes such as those mentioned above fuel the momentum. This has come as the science of prevention has evolved, and we have learned much about the factors leading to youth violence and poor developmental outcomes. Programs have been developed based upon research identifying specific risk and protective factors, and knowledge about how to successfully mitigate these factors, resulting in successful outcomes. Also, evaluation and statistical methods have evolved, bringing greater assurance of the effectiveness of programs. Such assurance has helped federal and local agencies, as well as private foundations, to be more proactive in promoting and/or requiring the use of evidence-based programs.

3. Creation of evidence-based registries

With an increased focus on what works, several federal and private registries have been created to help people sift through the findings and claims regarding evidence-based programs. These registries, while helpful, also create confusion among users because they vary widely in their focus and criteria for assessing effectiveness. Some registries were created to examine programs with specific outcomes. For instance, Blueprints, until recently, only examined outcomes of violence, delinquency, and substance use. Thus, a user looking for a program to address depression would previously not have found any listed. The terminology used to classify programs also varies, with some calling programs that meet the highest level of evidence “Model” and some labeling these programs as “Effective” or “Exemplary.” The criteria used to assess program effectiveness also vary from registry to registry. For instance, while one registry may call a program “Model,” another might list the same program as “Promising” or perhaps not list it at all because it does not meet the evidentiary standard. Registries also differ in whether or not they share information on programs reviewed but not meeting criteria.

Confusion around these differences often results in users assuming that if a program appears on a list that it has the same

level of evidence as any other program on the list. For instance, a survey of state education agency directors that administered the Safe and Drug Free Schools and Communities Act in 2004–2005 found that all states rely on federal lists to determine whether programs are evidence-based; however, they were willing to recommend any program that appeared on the lists, even if the evidence was weak (Hallfors, Pankratz, & Hartman, 2007).

Some help in understanding the evidence and ratings underlying these lists can be found on the Blueprints website which includes a “Matrix of Federal and Privately Rated Programs” that have been assessed by various registries (<http://blueprintsprograms.com/resources/matrix.pdf>). This matrix shows how a program has been rated across 5 agencies, including Blueprints, Coalition for Evidence-Based Policy, CrimeSolutions.gov, OJJDP Model Programs Guide, and SAMHSA-NREPP. Programs with consistently high ratings across several of the agencies should be ones that are given priority when searching for evidence-based programs to adopt. Users can also easily determine the top tier programs, regardless of the labeling used by each agency. The criteria used by each agency for assessing effectiveness are also provided.

4. Advantages of evidence-based programs

4.1. Assurance that the program works

There are numerous advantages in using evidence-based programs, but the biggest may be the assurance that the program works—the higher the scientific standard, the greater the assurance. Many people cling to the adage, “it’s better to do something than nothing.” Unfortunately, some things that are implemented with the best of intentions do harm to youth. Lipsey (1992) reported that approximately one-third (29%) of controlled evaluations of juvenile programs found negative effects. For instance, Scared Straight programs, which are currently being glamorized by a popular television network, have shown that the experimental groups reoffended between 1% and 30% more than the control groups (significance not provided) in seven of nine randomized studies reported on in a systematic review by Petrosino, Turpin-Petrosino, and Finckenaue (2000). The other two studies did not report group failure rates, and only one study reported positive effects (on new court intakes). Despite the best of intentions, this program led to more crime, showing that programs can not only fail, but may even do more harm than good.

Another example of a harmful program is the 21st Century Community Learning Centers, an afterschool program authorized by Congress in 1994. Nearly all centers offer recreational opportunities ranging from unstructured free time to organized sports. Programs also offer enrichment activities such as dance, drama, and music, as well as workshops on developmental topics such as building leadership skills and resolving conflicts with peers. The national evaluation found higher rates of negative behaviors (suspensions, disciplinary actions, calls to parents about bad behavior) in the treatment group, along with no differences in academic performance and mixed findings on developmental outcomes. The only positive outcome was that program youth felt safer than control youth (James-Burdumy, Dynarski, & Deke, 2007, 2008). This program grew quickly with an appropriation of \$40 million in fiscal year 1998 to \$1.49 billion in fiscal year 2014 (<http://www.afterschoolalliance.org/policy21stccclcfm>), despite evaluation evidence of its harmful effects. Investments in such ineffective and sometimes harmful programs are a waste of scarce violence prevention dollars and undermine public confidence in prevention science when the intended results are not achieved.

4.2. Cost-benefit data available

A second advantage is that cost-benefit data are often available for evidence-based programs, usually showing that the monetary benefits of implementing the program outweigh the costs. These data have been reported for various evidence-based programs by the Washington State Institute for Public Policy (WSIPP, 2013). For example WSIPP reports \$25.61 in monetary benefits per \$1 spent in implementing the LifeSkills Training program. Similarly, research at Pennsylvania State University reports \$25.72 in benefits from LifeSkills Training per \$1 spent, with an estimated \$16,160,000 in potential economic benefit statewide (Jones et al., 2008). WSIPP estimates that the implementation of a modest or aggressive portfolio of evidence-based programs in the State of Washington between 2008 and 2030 could save the taxpayers from \$1.9 to \$2.6 billion, respectively (Aos, Miller, & Drake, 2006). In the current decade of cost containment, showing that programs can actually generate savings is an important decision-making tool when deciding which investments in programs should be made.

4.3. Packaged/manualized materials available

Evidence-based programs offer packaged/manualized materials that explain what should be delivered to whom, when, where and how. The materials usually include protocols or checklists for monitoring implementation fidelity. In many of the evidence-based programs, quality assurance procedures are integral to the program and serve to improve the quality of implementation and the likelihood of results. Training and technical assistance are also available and are typically mandatory. An evidence-based program with packaged materials is usually much more time and cost-efficient than developing one's own program. Also program development should always be followed by evaluation, and this can be an extremely costly endeavor. An evidence-based program does not require evaluation in every new site during the adoption and early implementation stage, as its effectiveness has already been established. It is more important to monitor implementation fidelity to ensure that the program is being implemented as intended in order to ensure the best results. After the program has been established at a site, some form of evaluation may be warranted to ensure the program is having its intended effect at the site. It should be noted, however, that a program evaluation establishes effectiveness with a specific targeted population. A new evaluation is required when a program is implemented outside the population for which evidence exists.

5. Current use of evidence-based programs

The advantages of evidence-based programs in resolving the youth development issues faced by our nation seem unquestionable. Yet, the adoption of these programs has lagged behind the growth in the number of programs being made available. In mental health, the time lag between development of an evidence-based practice and its integration into routine practice is estimated to be 20 years (Hoagwood, 2003–2004; Institute of Medicine Committee on Quality of Healthcare in America, 2001). The implementation of evidence-based programs in school settings has grown over the years since national surveys first assessed their use, from 34.4% in 1999 to 42.6% in 2005 to 46.9% in 2008 (Ringwalt, Vincus, Hanley, Ennett, Bowling, & Haws, 2011; Ringwalt, Vincus, Hanley, Ennett, Bowling, & Rohrbach, 2009). Respondents in these surveys were also asked to identify which curriculum they used the most since most school districts use multiple curricula. Although nearly half the school districts used an evidence-based program in 2008, only 26% of schools used the

evidence-based program most frequently. Other studies have shown a similar pattern, reporting that a majority of schools use evidence-based prevention curricula, but they are rarely the most commonly used curricula (Pankratz & Hallfors, 2004). Additionally, Hallfors and Godette (2002) found that only 19% were implementing evidence-based curricula with fidelity. The use of evidence-based programs in other domains is even lower. Kumpfer and Alvarado (2003) estimate that in 2003 only 10% of family programs were evidence-based. Much more needs to be done to inform funders and practitioners of the benefits of evidence-based programs, the availability of programs that meet specific needs, the outcomes that can be achieved, and cost benefits that result. This has been the mission of Blueprints for Healthy Youth Development since 1996.

6. Blueprints for Healthy Youth Development

6.1. Background

Blueprints for Healthy Youth Development, formerly known as Blueprints for Violence Prevention, was one of the early leaders in the movement to identify evidence-based programs, promoting and using a rigorous standard. Blueprints, hosted at the University of Colorado Boulder—Center for the Study and Prevention of Violence, began in 1996 with several small grants and received major funding from the Office of Juvenile Justice and Delinquency Prevention beginning in 1997. The original focus was on the prevention of youth violence, delinquency, and substance use. In the latter half of 2010, with funding from the Annie E. Casey Foundation, the focus expanded to include youth programs to improve emotional and physical health and well-being, positive relationships, and academic success. The outcomes of interest involve more than preventing harmful behavior—they also involve positive behaviors and healthy development.

Blueprints identifies and recommends programs for children, youth and families that have undergone rigorous evaluations which have demonstrated strong evidence of effectiveness. Blueprints provides one of the highest standards in the field for quality programming.

6.2. Blueprints standard

The Blueprints' standards for recommending a program are among the most rigorous in the field. Many of the program registries ask for nominations and only review those studies submitted by the program developers. Thus, studies with null results may be omitted from the review process. Although Blueprints does review nominated programs, Blueprints additionally performs an exhaustive search of the literature on a monthly basis to find programs related to the outcomes of interest (delinquency, substance use, emotional and physical well-being, academic success, and positive relationships). A Blueprints program review typically considers *all* evaluations of a program. A comprehensive write-up of every program identified is completed. Each program write-up includes a description of the program, target audience, risk and protective factors, evaluation methodology, outcomes, generalizability, and limitations of each evaluation conducted for that program. A program then undergoes an internal review administered by Blueprints staff to determine if it might meet Blueprints criteria. Programs that pass this initial screening will undergo a second review conducted externally by the Blueprints Advisory Board of experts.

Blueprints considers four criteria for certifying a program:

- Evaluation quality: Studies must be of sufficient methodological quality to confidently attribute results to the program.

- **Intervention impact:** The preponderance of evidence from the high quality evaluations indicates significant positive change in intended outcomes that can be attributed to the program, and there is no evidence of harmful effects.
- **Intervention specificity:** The program description clearly identifies the outcome the program is designed to change, the specific risk and/or protective factors targeted to produce this change in outcome, the population for which it is intended, and how the components of the intervention work to produce this change.
- **Dissemination readiness:** The program is currently available for dissemination and has the necessary organizational capability, manuals, training, technical assistance and other support required for implementation with fidelity in communities and public service systems. Cost information and monitoring tools must also be available.

Blueprints programs must meet all four criteria. Programs are rated as either “Promising” or “Model.” Promising programs meet the minimum standard of effectiveness, requiring a minimum of (a) one high quality randomized controlled trial or (b) two high quality quasi-experimental evaluations. Model programs meet a higher standard and provide greater confidence in the program’s capacity to change behavior and developmental outcomes. Model programs require replication with a minimum of (a) two high quality randomized controlled trials or (b) one high quality randomized controlled trial plus one high quality quasi-experimental evaluation. There is also a requirement for sustained positive intervention impact for a minimum of 12 months after the program intervention ends.

To date, of more than 1300 programs assessed, 54 programs have qualified for Blueprints certification. While many programs with randomized or quasi-experimental designs have been evaluated, they often fail to meet the evaluation quality standard. [Appendix A](#) briefly describes the methodological elements that should be addressed in an evaluation. Some of the more common problems that prevent certification of a program include: (a) the failure to demonstrate baseline equivalence, (b) failure to determine if attrition differs by study condition, (c) randomizing subjects at one level (such as school) and conducting the analysis at a different level (such as individual), and (d) failure to follow and analyze all subjects as assigned to their original condition (intent to treat).

7. Blueprints website

Blueprints has developed a new website (<http://blueprintsprograms.com>) that highlights the model and promising programs in five outcome domains: problem behavior, emotional regulation, academic success, physical health and well-being, and positive relationships. Easy-to-use program searches allow users to match programs to identified needs. Searches can be run using criteria such as risk and protective factors, program outcomes achieved, type of program, or targeted population (i.e., age, gender, race/ethnicity), as well as several other factors. Each program listed on the website contains information on:

- description of the intervention,
- program goals,
- risk and protective factors,
- logic model (if available),
- outcomes achieved,
- training and technical assistance,
- contact information,
- program costs,
- funding strategies,
- cost-benefit information (if analyzed by Washington State Institute for Public Policy),

- full write-ups describing all evaluations of a program, including methodology, outcomes, and limitations.

The website includes links to needs assessment surveys for those who want to match a program to the needs identified in their schools or communities. The Blueprints website provides a brief description of the standards that various federal and private agencies use to rate programs, as well as a matrix of programs with the rating given by each of those agencies, allowing a comparison of ratings across agencies (<http://blueprintsprograms.com/resources>).

The information provided on the Blueprints website, as well as a national/international Blueprints Conference (<http://blueprints-conference.com>) held every two years, has led to greater awareness and use of evidence-based programs. The Blueprints website has served as a resource for governmental agencies, schools, foundations, and community human service organizations trying to make informed decisions about their investments in youth programs.

8. Lessons learned

Throughout the 17 plus years of the Blueprints project, the most important lesson that we have learned, and have struggled to impart, is that the standard for recognizing an evidence-based program must be high in order to maintain the public’s confidence. In the earliest years of the project, Blueprints identified a model program, Quantum Opportunities, based upon a multi-site evaluation (Hahn, 1994, 1995). This evaluation examined outcomes in each of the national sites participating in the study, and the demonstration appeared successful. Later, a large multi-site replication by the Department of Labor found only a few, largely inconsistent effects (Rodriguez-Planas, 2012), and some of the primary behavioral outcomes were negative at one of the replicating sites (unpublished report). This program had to be removed from the Blueprints list.

This happened again with another program, CASASTART, a comprehensive case management strategy for preventing drug use and delinquency for small groups of high-risk adolescents, ages 11–13, living in highly distressed neighborhoods. Although this program was only on the Blueprints list as promising, based upon one successful large-scale randomized trial conducted by the Urban Institute (Harrell, Cavanagh, Harmon, Koper, & Sridharan, 1997; Harrell, Cavanagh, & Sridharan, 1998), a later evaluation conducted by the Blueprints team found no effects, and some iatrogenic effects for girls (Mihalic, Huizinga, Ladika, Knight, & Dyer, 2011). This program was subsequently retracted by the program developers. However, many agencies were already using the program and were left with decisions of whether to continue or abandon the program. Some were in the midst of grants, while others were in the process of renewing grants that would support implementation. While most agencies ultimately decided to continue the use of the program, believing that they were having success with their youth, there was confusion as to how to handle the unexpected news.

These examples reinforce the reasoning for maintaining a high standard. We cannot afford to take programs to scale prior to rigorous testing. When a program has not been adequately tested and later evidence suggests that the program should be removed from a list, this shakes the confidence of the public. In the examples above, the original trials were conducted in a rigorous manner with random assignment, and later studies failed to replicate. This reinforces the need for replication as a part of the standard. There is always a chance of instances such as these happening, but to accept programs with an even lower standard than Blueprints could multiply such problems. The only

way to avoid these costly and confusing problems is to endorse standards of the highest quality, especially when taking programs to scale.

Blueprints maintains a rigorous standard; however, we suggest that as time goes by even our standard could be strengthened, especially by requiring *independent* replication. Currently, Blueprints model programs require replication; however, those studies are typically done by the program developers. There is accumulating evidence that programs evaluated by the program developers report effect sizes considerably larger than trials conducted by independent researchers (Eisner, 2009). This can be attributed to specific biases that can be inadvertently built into an evaluation at various steps in the process. Additionally, program developers evaluating their own programs provide a range of technical assistance resources that are often not available in real-world applications. At the current time, adoption of that standard would limit the number of model programs to less than a handful. We are hopeful that over time, researchers and funders will see the value of independent replication and more independent studies will be conducted.

9. Conclusion

Human services systems spend millions of dollars each year to address the problems that prevent youth from attaining developmental milestones and reaching their full potential. Unfortunately, even with the best of intentions, much of this work has not demonstrated the intended benefits, and some of the work done in public service agencies is conducted with little attention to behavioral outcomes. This has been changing, especially over the last decade, with an increased emphasis on accountability and promotion of evidence-based programs. However, public service systems and schools do not always know which programs would be best in meeting their needs. Often choices are made because of “hearsay” testimonial or other reasons not based on evidence. These systems need assistance in choosing programs that are effective and that represent a good match to the needs, programming, and staffing of the system.

Blueprints for Healthy Youth Development helps to meet those needs. Blueprints maintains a registry of programs that meet the highest standards of quality, and each program has been deemed ready for dissemination. The rigor underlying evaluations of Blueprints programs can assure users that they can make a real difference in preventing antisocial behavior and promoting positive behavior, academic success, emotional well-being, physical health and positive relationships.

The Blueprints website (<http://blueprintsprograms.com>) provides useful information on every model and promising program. There are easy-to-use program searches that help users find a program that matches their selection criteria. Users can search for a program by the outcomes they are interested in impacting, by a specific type of program, by risk and protective factors, and by the intended target of intervention (such as age, gender, race/ethnicity). Program information includes descriptions of the intervention, targeted audience, risk and protective factors, contacts, training and technical assistance, program costs, funding strategies, and cost-benefit data. There are also full descriptions of all studies conducted of a program, with information on methodology, outcomes, and limitations. The information is comprehensive, some of which cannot be found on other registries.

Blueprints will continue to promote the use of evidence-based programs and the use of high standards to identify those programs. Blueprints conducts a conference every two years with this as a major goal. Blueprints remains a leader in setting the standards for identifying evidence-based programs and for promoting the use of these programs nationally and internationally.

Appendix A. Checklist for Blueprint Program Evaluation and Write Up

Program Name _____

Primary criteria

___1. *Does the study have a high-quality design?* A randomized trial is ideal but two or more studies with quasi-experimental designs may be sufficient. Report on the use of randomization or the nature of the quasi-experimental design.

___2. *Does the study clearly describe the sample size at each stage of data gathering?* Report the number of subjects at each stage, including the N at baseline and the Ns and percentages of the baseline sample remaining at posttest and each follow-up.

___3. *Is measurement of the outcomes done independently from the delivery of the intervention?* Report on whether those doing the measurement are blind to condition. Subject self-reports are generally appropriate but may be subject to social desirability and demand bias.

___4. *Are the measures reliable and valid?* Report the information provided by the study (e.g., interrater reliability; Cronbach's alpha).

___5. *Are outcome measures general enough that they do not depend on the unique content of the intervention?* Report on measures that are unique to the program or not relevant outside the intervention.

___6. *Does the study use an intent-to-treat analysis?* The study should attempt to follow and analyze all subjects as assigned to their original condition.

___7. *Is the analysis done at the proper level?* Report on whether the analysis matches the level of the intervention (e.g., if schools are randomized, the analysis should compare schools, not persons, or use multilevel statistical methods that adjust for clustering).

___8. *Does the analysis control for baseline outcome measures?* Report on the use of change scores, baseline outcomes as covariates, or group-by-time interactions.

___9. *Does the analysis demonstrate baseline equivalence between conditions?* Report on whether a test was performed, nonequivalent findings, and potential adjustments.

___10. *Does the study demonstrate that attrition is below 5% or unrelated to group assignment, sociodemographic characteristics, and baseline measures of the outcomes?* Report on whether a test was performed, evidence of significant differential attrition, and potential adjustments.

___11. *Posttest: Are the results consistently significant – beyond what would be expected by chance – across multiple measures and statistical tests?* Report the number of significant results relative to the number of significance tests.

___12. *Are the outcomes free of any iatrogenic effects?* Report any significant effects that suggest program-related harm rather than help for subjects.

Secondary criteria

___13. *Does the study describe how the intervention is expected to produce change in the outcome?* Report the posited causal mechanisms.

___14. *Is the sample general?* Note if the sample is highly selective or unrepresentative, and describe the sociodemographic characteristics of the sample subjects.

___15. *Does the study provide quantitative measures of fidelity of implementation?* Describe.

___16. *Are effect sizes presented?* Report figures on Cohen's d, r, odds ratios, etc.

___17. Does the study examine mediators that help account for the relationship between the intervention and outcome? The mediating variables should measure risk and protective factors that are posited to translate the program into the outcomes.

Model criteria

___18. Long-term: Does the study demonstrate effects that last at least one year beyond the end of the intervention? Report on whether the study examined the sustained or long-term impact of a program with assessments following the posttest.

___19. Is the program studied replicated by a high-quality study?

Summary

___20. Should the program be recommended for review by the BP board? Sum up your view.

References

- Aos, S., Miller, M., & Drake, E. (2006). *Evidence-based public policy options to reduce future prison construction, criminal justice costs, and crime rates*. Olympia: Washington State Institute for Public Policy.
- Baron, J. (2012). Applying evidence to social programs. *Economix* Retrieved from <http://coalition4evidence.org/468-2/publications/> Published online 29.11.12.
- Baron, J. (2007). *Commentary: Making policy work: The lesson from medicine. Education Week* Retrieved from <http://coalition4evidence.org/468-2/publications/> Published online 23.5.07.
- Campbell, D. T., & Boruch, R. (1975). Making the case for random assignment to treatment by considering the alternatives: Six ways in which quasi-experimental evaluations in compensatory education tend to underestimate effects. In C. Bennet & A. Lundaine (Eds.), *Evaluation and experiments: Some critical issues in assessing social programs* (pp. 195–296). New York: Academic Press.
- Children's Services Council. (n.d.). *The journey to evidence-based programming. Children's Services Council of Palm Beach County*. Retrieved from <http://www.csepc.org>.
- Coalition for Evidence-Based Policy (2014). *What works in social policy? Findings from well-conducted randomized controlled trials*. Retrieved from <http://evidencebased-programs.org>
- Cook, T. D., Shadish, W. R., & Wong, V. C. (2008). Three conditions under which experiments and observational studies produce comparable causal estimates: New findings from within-study comparisons. *Journal of Policy Analysis and Management*, 27(4), 724–750.
- Eisner, M. (2009). No effects in independent prevention trials: Can we reject the cynical view? *Journal Experimental Criminology*, 5, 163–183.
- Elliott, D. S. (2013). Crime prevention and intervention over the life course. In C. Gibson & M. D. Krohn (Eds.), *Handbook of life-course criminology* (pp. 297–316). New York: Springer.
- Elliott, D. S., & Mihalic, S. (2004). *Blueprints for Violence Prevention*. Boulder, CO: University of Colorado, Institute of Behavioral Science, Center for the Study and Prevention of Violence.
- Greenwood, P. (2006). *Changing lives: Delinquency prevention as crime control policy*. Chicago: University of Chicago Press.
- Hahn, A. (1995). *Quantum Opportunities Program: A brief on the QOP pilot program*. Waltham, MA: Brandeis University, Center for Human Resources.
- Hahn, A. (1994). *Evaluation of the Quantum Opportunities Program (QUOP): Did the program work? A report on the post secondary outcomes and cost-effectiveness of the QUOP program (1989–1993)* Waltham, MA: Brandeis University, Heller Graduate School, Center for Human Resources.
- Hallfors, D., & Godette, D. (2002). Will the 'principles of effectiveness' improve prevention practice? Early findings from a diffusion study. *Health Education Research*, 17(4), 461–470.
- Hallfors, D., Pankratz, M., & Hartman, S. (2007). Does federal policy support the use of scientific evidence in school-based prevention programs? *Prevention Science*, 8, 75–81.
- Harrell, A. V., Cavanagh, S., Harmon, M. A., Koper, C. S., & Sridharan, S. (1997). *Impact of the Children at Risk program: Comprehensive final report* (Vol. 1). Washington, DC: The Urban Institute.
- Harrell, A. V., Cavanagh, S., & Sridharan, S. (1998). *Impact of the Children at Risk program: Comprehensive final report II: Executive Summary*. Washington DC: The Urban Institute.
- Hoagwood, K. (2003–2004). Evidence-based practice in child and adolescent mental health: Its meaning, application and limitations. *Emotional and Behavioral Disorders in Youth*, 4, 7–8.
- Institute of Medicine Committee on Quality of Healthcare in America (2001). *Crossing the quality chasm: A new health system for the 21st century*. Washington, DC: National Academy Press.
- Institute of Medicine (2008). *Knowing what works in health care: A road map for the nation*. Washington, DC: The National Academy Press.
- James-Burdumy, S., Dynarski, M., & Deke, J. (2007). When elementary schools stay open late: Results from the national evaluation of the 21st Century Community Learning Centers Program. *Educational Evaluation and Policy Analysis*, 29(4), 296–318.
- James-Burdumy, S., Dynarski, M., & Deke, J. (2008). After-school program effects on behavior: Results from the 21st Century Community Learning Centers Program national evaluation. *Economic Inquiry*, 46(1), 13–18.
- Jones, D., Bumbarger, B. K., Greenberg, M. T., Greenwood, P., & Kyler, S. (2008). *The economic return on PCCD's investment in research-based programs: A cost-benefit assessment of delinquency prevention in Pennsylvania*. The Prevention Research Center for the Promotion of Human Development, Pennsylvania State University.
- Kumpfer, K. L., & Alvarado, R. (2003). Family-strengthening approaches for the prevention of youth problem behaviors. *American Psychologist*, 58, 457–465.
- Lipsey (1992). The effect of treatment on juvenile offenders: Results from meta-analysis. In F. Losel, D. Bender, & B. Gliesener (Eds.), *Psychology and law: Interpersonal perspectives* (pp. 131–143). Berlin: deGruyter.
- Martinson, R. (1974, June). What works? Questions and answers about prison reform. *Public Interest*, 22–54.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional & Behavioral Disorders in Youth*, 4, 83–104.
- Mihalic, S., Huizinga, D., Ladika, A., Knight, K., & Dyer, C. (2011). *CasaStart final report*. Boulder, CO: University of Colorado, Center for the Study and Prevention of Violence.
- Office of Management Budget (2001). *The president's management agenda*. Washington, DC: US Government Printing Office.
- Office of Management Budget (2002). *The president's management agenda*. Washington, DC: US Government Printing Office.
- Pankratz, M. M., & Hallfors, D. D. (2004). Implementing evidence-based substance use prevention curricula in North Carolina public school districts. *Journal of School Health*, 74(9), 353–358.
- Petrosino, A., Turpin-Petrosino, C., & Finckenauer, J. O. (2000). Well-meaning programs can have harmful effects! Lessons from experiments of programs such as scared straight *Crime and Delinquency*, 46(3), 354–379.
- Redirection (2012, October). *Redirection quarterly update report, Florida's commitment to helping at-risk youth*. Retrieved from www.evidencebasedassociates.com
- Regnery, A. S. (1985). Getting away with murder. *Policy Review*, 34, 1–4.
- Ringwalt, C., Vincus, A. A., Hanley, S., Ennett, S. T., Bowling, M., & Haws, S. (2011). The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2008. *Prevention Science*, 12, 63–69.
- Ringwalt, C., Vincus, A. A., Hanley, S., Ennett, S. T., Bowling, M., & Rohrbach, L. A. (2009). The prevalence of evidence-based drug use prevention curricula in U.S. middle schools in 2005. *Prevention Science*, 10, 33–40.
- Rodriguez-Planas, N. (2012). Longer-term impacts of mentoring, educational services, and learning incentives: Evidence from a randomized trial in the United States. *American Economic Journal: Applied Economics*, 4(4), 121–139.
- Romig, D. (1999). *Justice for our children: An examination of juvenile delinquency rehabilitation programs* (3rd ed.). Troy, MI: Performance Resource Press.
- Sawhill, I. V., & Baron, J. (2010). Federal programs for youth: More of the same won't work. *Youth Today* Retrieved from <http://viewpoint-Essay-Sawhill-Baron-Youth-Today-May-2010> Published online 1.5.10.
- Sechrest, L., White, S., & Brown, E. (1979). *Rehabilitation of criminal offenders: Problems and prospects*. Washington, DC: National Academy of Sciences.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation*. Newbury Park, CA: Sage Publications Inc.
- Sherman, L., Farrington, D., Welsh, B., & McKenzie, D. (Eds.). (2002). *Evidence-based crime prevention*. New York: Routledge.
- Society for Prevention Research. (n.d.). *Standards of evidence: Criteria for efficacy, effectiveness and dissemination*. Retrieved from <http://preventionresearch.org>.
- Taylor, N. J., Conner, M., & Lawton, R. (2012). The impact of theory on the effectiveness of worksite physical activity interventions: A meta-analysis and meta-regression. *Health Psychology Review*, 6(1), 33–73.
- Webb, T. L., Joseph, J., Yardley, L., & Michie, S. (2010). Using the internet to promote health behaviour change: A meta-analysis of the impact of theoretical basis, use of behavior change techniques, and mode of delivery on efficacy. *Journal of Medical Internet Research*, 12(1), e1.
- WSIPP (2013). *Benefit-cost results*. Retrieved from <http://wsipp.wa.gov/BenefitCost>

Sharon F. Mihalic, M.A. has been a researcher at the University of Colorado Boulder for 24 years. She has helped to facilitate multiple facets of work involved in conducting a major longitudinal, national sample to collect data on juvenile delinquency. Research, using this survey, includes articles in the areas of marital violence, drug use, and the effects of adolescent employment on delinquency. During the last 18 years, her major role at the Center for the Study and Prevention of Violence at the University has been as the Director of the Blueprints for Healthy Youth Development initiative (formerly Blueprints for Violence Prevention). She has examined the evaluations of numerous violence prevention programs and has had major input into the selection of the Blueprints programs. She is a co-author or contributing author on the twelve Blueprints books, as well as the volume editor of each book, and the co-editor of the Blueprints series. She provided the direction and management for two past Blueprints dissemination projects to replicate the Blueprints programs in multiple sites nationwide. She was the PI on a project funded by the Robert Wood Johnson Foundation to replicate and evaluate two of the Blueprints promising programs. She is a Co-PI on a corporate-funded grant that disseminates and conducts process evaluation for the LifeSkills Training program in 15 states. Research from the Blueprints initiative includes seven articles, two bulletins, and a monograph on the factors associated with implementation success. She received the "Science to Practice" Award from the Society for Prevention Research in 2008.

Delbert S. Elliott is the Acting Director of the Program on Problem Behavior and the Founding Director of The Center for the Study and Prevention of Violence in the Institute of Behavioral Science at the University of Colorado Boulder. He is also a Distinguished Professor Emeritus and currently a Research Professor in IBS. He directed a number of national longitudinal studies and randomized control trials of violence, drug and delinquency prevention programs, including the National Youth Survey, the longest study of antisocial behavior, delinquency, violent behavior and drug use in a representative national panel of adolescents and young adults in the United States. Del's books include *Delinquency and Dropout* (1974); *The Social Psychology of Runaway* (1978); *Explaining Delinquency and Drug Use* (1985); *Multiple Problem Youth: Delinquency, Drugs and Mental Health Problems* (1989); *Violence in American Schools* (1998) and *Good Kids from Bad Neighborhoods* (2006). Del is the founder of the Blueprints for Violence Prevention Initiative and the Senior Editor of *Blueprints for Violence Prevention*, a series of monographs describing model violence prevention programs. Del served as Chair of the Criminal and Violent Behavior Review Committee (NIMH) and is a past President and Fellow of the American Society of Criminology and former

member of the Advisory Board for the Center for Injury Prevention and Control, Centers for Disease Control. In 1995 he received the Edwin H. Sutherland Award for outstanding contributions to the field of Criminology from the American Society of Criminology; in 1998 he received an Outstanding Achievement Award from the Office of Juvenile Justice and Delinquency Prevention (U.S. Department of Justice). In 2000 he received the Paul Tappan Award for Fundamental Contributions to Knowledge of Youthful Offenders and Domestic Violence by the Western Society of Criminology and the Science to Practice Award from the Society for Prevention Research for outstanding efforts to identify and disseminate empirically supported interventions to prevent crime and delinquency. He was the Senior Science Editor for *Youth Violence: A Report of the Surgeon General*, 2001. In 2001 he received the Public Health Service Medallion for Distinguished Service awarded by Dr. David Satcher, U.S. Surgeon General. In November of 2003 he received the August Vollmer Award from the American Society of Criminology, in 2005 he was elected a Fellow of the Academy of Experimental Criminology, and in 2008 he was elected a Fellow of the Society for the Advancement of Evidenced Based Practices.