

DEBATE

EVIDENCE-BASED APPROACHES TO IMPROVING JUVENILE
JUSTICE PROGRAMMING

Evidence-based juvenile justice programs and practices: A critical review

Delbert S. Elliott¹  | Pamela R. Buckley¹  |
Denise C. Gottfredson²  | J. David Hawkins³ | Patrick H. Tolan⁴ 

¹ University of Colorado Boulder

² University of Maryland

³ University of Washington

⁴ University of Virginia

Correspondence

Delbert S. Elliott, Institute of Behavioral Science, University of Colorado Boulder, 483 UCB, Boulder, CO 80309.

Email: elliott1100@comcast.net

This work was supported by a grant from Arnold Ventures. The views and opinions expressed here are solely those of the authors and may not be attributed to the Arnold Ventures Foundation.

Funding information

Arnold Ventures Foundation

Abstract

There is growing critical commentary and debate about the relative effectiveness of individual program and generic practice approaches to identifying evidence-based interventions and their impact on the operation of the juvenile justice system. The central issue is whether both of these approaches to identifying evidence-based interventions provide a valid and reliable guide to improving juvenile justice programming and, if so, what are the relative advantages and disadvantages of each? From a public policy perspective, should we be investing more heavily in one or the other, or treating them as effective complementary approaches and encourage both? We address each of these questions and offer some suggestions for improving the effectiveness of each approach.

KEYWORDS

Certification Standards, Cost-Benefit, Delinquency Prevention, Evidence-Based, Fidelity, Practices, Programs

Intervention science is the study of the development, testing, dissemination and implementation of effective treatments and prevention models in service of at-risk and vulnerable groups, and focuses on three basic types of interventions: programs, practices, and policies (Institute of Medicine, 2015; Springer & Phillips, 2007; Walker, Lyon, Aos, & Trupin, 2017). *Programs* involve

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2020 The Authors. *Criminology & Public Policy* published by Wiley Periodicals LLC on behalf of American Society of Criminology

a coherent package of activities with defined delivery protocols, implementation manuals, training, and technical assistance that implement an identified logic model, i.e., a particular change strategy targeting identifiable risk or protective factors that are theoretically linked to a specific outcome, like recidivism. There is less consensus regarding the definition of *practices*, which typically involve generic types of programs or strategies that have some common elements or core components but are more flexible than manualized programs as they do not necessarily involve the same detailed package of prescribed activities that characterize programs. *Policies* involve formal regulations or laws that apply uniformly to general populations (Elliott & Fagan, 2017; Substance Abuse and Mental Health Services Administration, 2018; Walker et al., 2017). With respect to current juvenile justice prevention and treatment, interventions demonstrated by research to be effective, hereafter referred to as evidence-based interventions, are primarily programs and practices. Currently, almost no juvenile justice policies are considered evidence-based (Elliott, 2017; Elliott & Fagan, 2017). Our focus in this essay is on comparing the utility of *evidence-based programs and practices for guiding juvenile justice intervention*.

Critical commentary and related debate about the relative effectiveness of these two approaches to identifying evidence-based interventions and their impact on the operation of the juvenile justice system have intensified (Embry & Biglan, 2008; Gorman et al., 2007; Greenwood, 2018; Lipsey, 2018; Mullen & Streiner, 2004; Rehuher, Hiramatsu, & Helm, 2008; Welsh, Rocque, & Greenwood, 2014). Several critical research and public policy questions are at the center of the debate: What criteria should be used to evaluate the relative effectiveness of these two approaches? How effective is each approach based on these criteria? What evidence is there that each can be implemented with fidelity and to scale? How usable by practitioners is each? Based on such assessment, should we be investing more heavily in one than the other? How might we improve the effectiveness of each? In this essay we address these questions to inform decisions about the relative viability of each approach for directing juvenile justice system interventions.

1 | EVALUATING THE EFFECTIVENESS OF PROGRAM AND PRACTICE APPROACHES

Researchers use a basic set of standards for evaluating the effectiveness of program and practice approaches. First, how extensive and robust is the evidence base for each? This includes questions about the quality, quantity, and level of evidence established and the scientific standards used in each approach for determining whether a program or practice is “evidence-based.”

Second, how extensive and sound is the research evidence for the claim that specific programs or a set of practices are effective in reducing juvenile reoffending when scaled up and implemented as routine practice in the juvenile justice system?

Third, what is the expected impact of each approach on the population of juvenile offenders in the justice system? How soundly can that impact be inferred? Should programs or practices with larger effect sizes or serving a larger segment of offenders be given funding priority? What policy implications follow from the findings about each approach?

Fourth, what does the existing evidence say about likely juvenile justice system adoption rates and factors influencing adoption decisions? These include factors like fidelity requirements and differences in costs and benefits.

Clearly, additional criteria might be considered to compare the strengths and weaknesses of these two approaches. However, this set addresses the most fundamental considerations that have been emphasized in prior commentary (Embry & Biglan, 2008; Greenwood, 2018; Lipsey, 2018; Mullen & Streiner, 2004).

1.1 | Evidence base for programs and practices

1.1.1 | Program approach: Registries

Three online, regularly maintained registries of evidence-based programs serve juvenile offenders and those at risk of delinquent offending: 1) the Office of Justice Programs' (OJP's) *CrimeSolutions* (<http://www.crimesolutions.gov/programs.aspx>); 2) the Office of Juvenile Justice and Delinquency Prevention's (OJJDP's) *Model Programs Guide* (<http://www.ojjdp.gov/mpg>); and 3) *Blueprints for Healthy Youth Development* (formerly *Blueprints for Violence Prevention*), a privately funded registry housed at the University of Colorado Boulder (<http://www.blueprintsprograms.org>). All three registries rate programs using scientific standards that require the research design implemented to establish causal inference sound enough to rule out alternative explanations for findings (i.e., sources of systematic bias). Interventions are then ranked as being "promising" or "effective" (OJP and OJJDP) or "promising" or "model" (Blueprints) based on the strength and extent of evaluation findings. Our concern in this essay is with programs rated as "effective" or "model," an evidence-level rating that the Blueprints registry specifically notes is required for scale-up, and that CrimeSolutions and the Model Program Guide define as programs that have been demonstrated to "work."

The Blueprints registry standard for a model rating requires: 1) a theoretical rationale/logic model, 2) one or more high quality randomized controlled designs (RCTs), 3) an experimental replication, 4) sustainability of effects for at least one year post intervention, 5) no evidence of iatrogenic effects, and 6) the organizational capacity to provide materials, training, and information for potential users to adopt and implement the model program with fidelity. Blueprints also has a "model+" rating, which requires at least one trial by an evaluation team independent of the program developer and his or her colleagues. In sum, the logic is that for Blueprints to consider any program as ready to be relied on in practice or for scaling up, there should be a clearly specified theoretical rationale, reliable evidence of positive, sustained impact based on a sound experimental research design, and at least one replication of positive findings.

CrimeSolutions and the Model Program Guide registries use the same evidence standard to certify a program as effective. The standard is based on a complex assessment of each experimental trial on: 1) the theoretical/conceptual framework, 2) the type of research design (an RCT or quasi-experimental design [QED], 3) the strength of the outcome, and 4) the fidelity of implementation. Effective interventions must have demonstrated positive effects in at least one high quality RCT or QED trial, with no evidence of an iatrogenic effect. This standard, which does not require an RCT or replication for certification as an effective program, is a lower evidentiary standard than used by Blueprints for either a model or promising rating, (see Fagan & Buchanan, 2016).

Extensive searches of existing evaluation studies find only a few model or effective programs specifically designed for juvenile justice system offenders that meet criteria for listing on these registries—four on the Blueprints registry and 10 on the Justice Department registries. If we include programs that impact delinquent offenders outside the juvenile justice system, the number of model or effective programs increases substantially. However, in this essay we focus on those programs specifically intended to serve offenders in the juvenile justice system.

The fact that relatively few programs qualify for inclusion on these registry lists has led to the criticism that evidence-based programs in juvenile justice are very limited and unable to cover the full range of juvenile justice service needs (Embry & Biglan, 2008; Lipsey, 2018). Given this limitation, Lipsey (2018) suggests that juvenile justice programming should focus on generic practices rather than on the evidence-based programs approach taken by these registries. The pertinent

question here is whether these few programs have a significant demonstrated impact on reoffending rates and how this compares to the demonstrated impact of evidence-based practices, which currently offer a broader range of services. Further, assuming both have a significant impact on reoffending rates, are there other advantages that favor the adoption of one over the other, or is the best policy to encourage both approaches? We will consider this specific issue later in the essay.

Blueprints model programs

The four Blueprints model programs serving juvenile justice youth are: Multisystemic Therapy (MST); Functional Family Therapy (FFT); Treatment Foster Care Oregon (TFCO); and MST-PSB (MST for Problem Sexual Offenders). MST, a delinquency prevention program to enhance parenting skills and provide intensive family therapy, has an evidence base that (to date) consists of 74 published studies, including 11 randomized controlled trials (RCTs) involving justice system juvenile offenders and seven involving substance abuse and delinquent behavior outside the justice system. Seven of the offender RCT studies were conducted by an independent evaluator (the criterion for a model+ rating by Blueprints). A specialized version of MST for sexual offenders (MST-PSB) has been implemented in three additional RCTs. FFT, a short-term family therapy and juvenile diversion program, has been evaluated in over 20 studies including 13 RCTs, five of which were independent evaluations. TFCO, a therapeutic foster care program, has 10 evaluation studies with six RCTs (<http://www.blueprintsprograms.org>; <http://www.ftllc.com/about-fft-training/fft-research.html>; Henggeler & Schoenwald, 2011; MST Services, 2019). In addition, the data base for each of these programs includes studies of mediation effects, which provide additional support for the causal mechanism upon which the program is based (<http://www.blueprintsprograms.org>). These model programs have a deep and broad evaluation evidence base documenting their effectiveness.

There is also extensive evidence for program effectiveness with scale-up for these programs. State-level initiatives have shown that they are scalable and when well implemented they yield substantial benefits. Washington was the first state to adopt and implement model programs statewide. In 1997, Washington passed the Community Juvenile Justice Act requiring the Washington State Institute for Public Policy (WSIPP) to evaluate the experimental evidence for delinquency prevention programs and identify effective interventions. WSIPP identified four programs, two of which were FFT and MST. Under the provisions of this Act, the state's juvenile courts were required to use one of these four programs for court-ordered placement and to eliminate funding for any existing program found to be ineffective in reducing recidivism. The statewide impact of this initiative was a 10% reduction in recidivism (Barnoski, 2004; Drake, Aos, & Miller, 2009).

Project Redirection in Florida represents another statewide adoption and implementation of model programs. The initiative first implemented FFT and MST in five judicial circuits. It served youth who violated their probation by committing misdemeanors or technical violations and were diverted to one of these model community programs or returned to a residential correctional setting. The reductions in recidivism and cost savings for participants in these programs were substantial, and the success of this initial project led to its expansion to 18 of the state's 20 judicial circuits. A 2010–2011 evaluation of this initiative found an 8% reduction in rearrests and a 24% reduction in felony rearrests for youths in the programs compared to youths returned to traditional state residential facilities. Reconviction rates were 38% lower and recommitment rates were 33% lower. This initiative with model programs served over 1,000 youth annually (The Justice Research Center, 2013). This intervention was awarded the Prudential-Davis Productivity Award for innovations in case processing and the 2008 Science and Service Award from the Substance Abuse and Mental Health Services Administration for its

success in helping communities adopt and implement effective programs (Elliott & Fagan, 2017). The initiative is now being implemented in Georgia with two Blueprint programs: FFT and a promising Blueprints program, The Strengthening Families 10–14 Program.

A third state, Pennsylvania, implemented a set of seven Blueprint model programs statewide. The Research-Based Programs Initiative achieved significant reductions in delinquency, violence, and drug and alcohol use (Jones et al., 2008). In the three years from fiscal year (FY) 2011/2012 through 2013/2014, MST served more than 4,500 youth. Forty percent were referrals from the juvenile justice system, and the reductions in recidivism saved the state an estimated \$27.3 million (EPISCenter, 2014). During this same period, FFT served 3,324 youth, and about 42% were referrals from the juvenile justice system, saving the state an estimated \$30.7 million (EPISCenter, 2015).

There is also evidence of larger scale implementation viability. In 2012, MST was selected by Chile's government for implementation with high-risk juvenile offenders and given three years to demonstrate its effectiveness. By the end of this period, MST had established 14 sites and served 2,574 youth. Currently MST is being used in 36 sites across Chile, has served over 10,500 youth with an average saving of \$11,000 per youth, and has achieved a 43% reduction in recidivism compared to Chile's national average in 2010 (MST Services, 2019).

There are few evidence-based Blueprint model juvenile justice programs, but the evidence base for those that have been identified is extensive. These programs have demonstrated replicated benefits, that they can be scaled up with fidelity, that they can be implemented as routine practice in the juvenile justice system, and that they achieve positive system-level effects.

CrimeSolutions and model program guide programs

The evidence base for the six effective programs uniquely listed in the OJP and OJJDP registries is more limited. In large part, this smaller number results from the Department of Justice's policy of including no more than three studies in their expert review even if more than three studies are found in the literature and from the lower evidence standard used by their registries. Two of the six effective programs (the Adolescent Diversion Project and Multisystemic Therapy-Substance Abuse) have at least two evaluation trials with one being an RCT. Three programs have only a single QED evaluation trial (Brooklyn Treatment Court, Juveniles Breaking the Cycle, and Project BUILD), and one, Aggression Replacement Training (ART), has been evaluated with two QEDs involving juvenile justice system offenders.

To our knowledge, little evidence is available on the scale-up of these programs, but some of them may have substantial levels of implementation. WSIPP recently conducted an evaluation of ART, which was one of the programs selected for the Washington statewide implementation under the Community Juvenile Justice Act discussed earlier. This evaluation involved the implementation of ART in 21 courts, and found iatrogenic effects for the program Knoth, Wanner, & He, 2019).

This difference in evidence for effective scale-up of programs certified by different registries highlights the importance of having high methodological standards when determining which programs are effective and the importance of replication across all methodologically adequate studies. It is clear that as criteria for certification vary, the evidence base for model and effective programs (hereafter evidence-based programs) will be inconsistent. The lack of consensus on the scientific standard for certifying programs as effective is an important issue that we will return to later.

1.1.2 | Practice approach

The evidence for certified practices is based primarily upon meta-analyses of sets of individual program evaluations that have some common feature; involve an experimental or quasi-experimental evaluation design demonstrating a positive relationship between the intervention and the outcome; and provide an estimated average effect size for the practice. The actual range in the number of programs included in a meta-analysis of a practice can range from a minimum of two (CrimeSolutions) to more than 50 (Lipsey, 2009). Because of this methodology, it is difficult to determine the breadth or depth of research for a given practice. Each review may use different search strategies and inclusion criteria, different techniques for addressing publication bias, and different meta-analytic techniques for handling dependencies. In addition, reviews may use different methodological requirements for primary studies or a different set of program categories, or categorize a given intervention differently. Also, as categories used become more general, more studies will be considered as addressing the validity of a given category. Thus, while it is likely that the number of studies included in the evidence base for some practice categories would be larger than that for many evidence-based (model or effective) programs, it is not clear that there are generally more sound studies of clearly defined practices than of evidence-based programs.

CrimeSolutions is the only registry currently listing evidence-based practices determined by meta-analysis. It currently lists 28, such as “disorder policing,” “contingency management interventions for substance use disorders,” and “dropout prevention programs.” All published and reported meta-analyses found in the literature search are included in the CrimeSolutions’ review. The standard for rating a practice as effective involves an assessment of the *overall quality* of each meta-analysis included in the evidence base for that practice and an assessment of the *internal validity* of outcomes in each meta-analysis, referring to the extent to which changes in the outcome can be attributed to the intervention practice. In rendering this judgement, priority is given to the quality of RCTs in the estimate of the overall practice effect size. For each practice rated as effective, the practice goals, target population, key components, practice theoretical rationale, and the number of programs included in each meta-analysis are described, along with the average effect size for the practice.

Only one listed evidence-based practice in CrimeSolutions to date specifically focuses on serving offenders in the juvenile justice system: *Treatment in Secure Corrections for Serious Juvenile Offenders*. The evidence base for this practice involved a single meta-analysis of 17 evaluation studies (six RCTs and 11 QEDs) with an average effect size odds ratio of 1.35.¹ While there are no other online, maintained registries of practices, there are other sources for meta-analyses of juvenile justice practices, such as the Campbell Collaboration (www.campbellcollaboration.org). However, Campbell does not render any rating of effectiveness based on an explicit scientific standard, nor is there standardization about how evaluation studies are categorized to determine practices. Currently, various meta-analyses vary greatly in how they determine which studies to include and how categories of practices are defined and differentiated, and in the validity of benchmarks for identifying a given practice as effective.

What do we know about the effectiveness of evidence-based practices when scaled up? To our knowledge, except for the Standardized Program Evaluation Protocol (SPEP), discussed next, no one has attempted a systematic scale-up of evidence-based juvenile justice practices. A critical question is how does an *average effect size for a generic practice* get applied to juvenile justice programming or influence the adoption, adaptation, or termination of a juvenile justice system intervention?

A fundamental challenge is how to determine what constitutes implementing a practice identified as having a significant average effect. One option would be to choose an individual program identified as belonging to a particular practice set that fits local needs and replicate it. However, the selected individual program cannot claim evidence-based status based on the practice's average effect size. Such a claim might be warranted if the practice were defined by a specific set of core components with an established causal link to the outcome. However, we know of no such practice definitions for juvenile justice interventions. A second option might be to create a new program that addresses unique local needs and resources by incorporating the core components or logic model common to all programs in the set. This would be a reasonable option, except that almost no existing research identifies the necessary and sufficient components of practices, and no meta-analysis we could find defines juvenile justice practice types based on shared core components or a logic model. A third option might be to compare programs currently operating in the juvenile justice system with programs included in an effective practice and adapting their programs to conform more closely to the core characteristics of the practice. But this option has the same limitations as the former option and depends on research demonstrating that critical components have been extracted and distinguished from noncontributory components of a tested program of this type. Finally, a fourth option is to consider terminating any program that is of a practice type found to be, on average, ineffective. Again, this could be a serious mistake, as it assumes the average effect size can be generalized to all individual programs in the practice. In each case, these strategies rest on an assumption that some common underlying causal mechanism among programs in a practice accounts for how the shared components that define the practice work to change targeted outcomes. Given current meta-analytic practice, this is an unwarranted assumption.

2 | STANDARDIZED PROGRAM EVALUATION PROTOCOL (SPEP)

The Standardized Program Evaluation Protocol (SPEP) is a system that aims to adapt juvenile justice programs currently in use to conform more closely to a set of practice characteristics shown to be related to program effectiveness. Lipsey (2009) identified a set of juvenile justice practices in a meta-analysis of 548 evaluation studies of juvenile justice programs conducted between 1958 and 2002. This extensive database provided the opportunity to examine trends across and variations among a wide range of programs implemented in the juvenile justice system before 2003 that had produced some estimate of evaluation effects.

To understand how the type of intervention approach might relate to effects, Lipsey aggregated individual evaluations into broad generic types of programs: family-based therapy, cognitive behavioral therapy, counseling, skill development, disciplinary interventions, deterrence-focused interventions, and surveillance-focused interventions. Effect sizes were then estimated for each type of practice on recidivism. He found small to moderate positive effects for most of his practice types and labeled this set as "therapeutic" practices. He also found several practices that were ineffective or harmful: discipline-oriented practices, deterrence practices, and surveillance practices. Lipsey subsequently applied a regression analysis to determine what program features were associated with lower recidivism rates in therapeutic practices and found four characteristics: 1) the risk level of offenders in the program, 2) the type of therapeutic programming being implemented, 3) the duration of service delivered, and 4) the quality of the intervention delivered. Based on this set of practices and the characteristics associated with recidivism, Lipsey developed an

intervention called the Standardized Program Evaluation Protocol (SPEP), which he promotes as an evidence-based practice intervention (Lipsey, 2018).

SPEP is a fidelity tool, assessing compliance to the set of four characteristics of effective therapeutic practices found in Lipsey's (2009) meta-analysis. Lipsey (2018) describes it as an assessment or measurement scheme based on program characteristics predictive of favorable recidivism effects. In essence, these four practice characteristics are treated as core components in an evidence-based intervention. SPEP is offered to juvenile justice systems as a set of evidence-based practice guidelines for improving the effectiveness of their current programming with the claim that increasing compliance levels across these four characteristics (i.e., their SPEP score) will increase their "expected effectiveness" (Lipsey, Howell, Kelly, Chapman, & Carver, 2010).

3 | META-ANALYSIS QUALITY AS A FACTOR IN PRACTICE RATINGS

None of the practices identified in Lipsey's meta-analysis is rated as an effective practice in the CrimeSolutions registry, although several are listed as promising practices. The validity of the meta-analysis upon which SPEP is based (Lipsey, 2009) is threatened by two problems: the variability in the quality of the evaluation studies included, and questions of the internal validity of the identified practices, the two criteria used by CrimeSolutions for evaluating meta-analyses. A large portion of the evaluations included in Lipsey's meta-analysis involved quasi-experimental designs (QEDs) with unmatched control groups. Further, of those matched, a substantial number were matched only on demographic characteristics and not on pretest measures of desired outcomes. Moreover, many included studies had small sample sizes, focused on nonserious offenders, did not document program content, logic model, or fidelity, or relied on a single, short-term follow-up to calculate effects. Lipsey himself described the general body of evaluation research prior to 1988, a major portion of the selection period for his meta-analysis, as "frankly horrid" (Lipsey, 1988, p. 6).

This reliance on low-quality QED studies to assess effectiveness of constructed practices raises concerns. Quasi-experiments can produce unbiased causal inference, but only if they meet the strong ignorability assumption (Rosenbaum & Rubin, 1984). This requires that the covariates used in a study capture the true process of selection into treatment that is correlated with the study outcome. However, in actual research practice with QEDs, only regression discontinuity designs have been shown to demonstrably meet this condition (Chaplin et al., 2018; Cook, Shadish, & Wong, 2008). It is unlikely that the QEDs in the set of studies included in Lipsey's meta-analysis met this assumption.

Further, the available evidence indicates that the inclusion of QEDs in meta-analyses provides inflated estimates of effect size and that there is no method to accurately estimate the extent or magnitude of these biases (Cheung & Slavin, 2015). Small sample sizes and weak internal validity also tend to inflate effect sizes (Cheung & Slavin, 2015; Coyne, Thombs, & Hagedoorn, 2010; Kraemer, Gardner, Brooks, & Yesavage, 1998; Weisburd, Lum, & Petrosino, 2001; Welsh, Peel, Farrington, Elffers, & Braga, 2011). These threats to internal validity cannot be rendered minor by statistical controls or by including many studies. Basically, QEDs provide weak evidence in determining effectiveness. The designation of a program or practice as evidence-based requires more definitive evidence like that provided by RCTs (National Research Council and Institute of Medicine, 2009).

4 | INTERNAL VALIDITY AS A FACTOR IN PRACTICE RATINGS

Campbell and Stanley (1963) view internal validity as the *sine qua non* of good evaluation research, highlighting the priority they give to causal inference. Does the evaluation provide compelling evidence that the intervention produced the observed effect in the outcome? Given the way practices are constructed in Lipsey's meta-analyses, it is unclear what causal mechanism is involved and what program components or characteristics account for the observed impact. Lipsey's (2009) practice taxonomy is not based on a program's core components, logic model, or theoretical rationale, but rather on broad philosophies of interventions developed "inductively." As a result, there may be substantial heterogeneity within each of his generic types of practice in core components, logic models, and possible causal mechanisms, as well as program content, population included, procedures, and activities. Heterogeneity is also reflected in the greater variation in effect sizes within practice types than between them (Lipsey, 2009). Such variation in what is being identified as a distinct practice raises questions about the appropriateness of using the average effect size to characterize the effectiveness of the practice (Durlak & Lipsey, 1991), the construct validity of Lipsey's taxonomy of generic practices, and any causal interpretation of the relationship between the intervention and recidivism.

Moreover, the set of evaluations included in each practice category incorporates interventions whose evaluation(s) indicated their estimated effect was negative. For example, in the family therapy practice, 14% of reported effects, excluding those for FFT and MST, were negative (intervention fared worse than controls). Nearly 10% of cognitive behavioral therapy study evaluations found negative effects (Lipsey, 2014; Lipsey et al., 2010). In some cases, these were interventions for which the *only* evaluation of their effectiveness found negative effects. If these are reliable and statistically significant effects, Lipsey's certified practices could include specific interventions demonstrated to be harmful. There is a potential for promoting harmful interventions when a practice with high heterogeneity in specific intervention activities or negative results is certified as a sound practice. Lipsey et al. (2010) acknowledge this problem,² but there is no indication that it is addressed in the typology of practices or implementation of SPEP.

In the evidence-based program approach utilized among the pertinent registries, finding a harmful effect for a program in a well-conducted trial disqualifies that program, even if the average effect size for the set of that program's evaluations is positive. The standards for certifying practices as evidence-based need some rule or guideline for dealing with interventions included in a type of practice that are found to be harmful.

While the formation of generic practice types identified by meta-analyses such as Lipsey (2009) may provide some heuristic utility for suggesting promising strategies and approaches for interventions in juvenile justice, as currently formulated they do not warrant being rated as evidence-based interventions that can be scaled up for use in the juvenile justice system. This is because the evidence for SPEP is limited to correlational evidence.³ With respect to the development of SPEP, Lipsey's regression analysis identifying the four program characteristics was an atheoretical exploratory analysis, most appropriately interpreted as suggesting hypotheses to be tested. It was not a confirmatory analysis. These regression findings do not provide any valid causal interpretation of the relationship between SPEP and recidivism (Jaffee, Strait, & Odgers, 2012).

With regard to the evidence for SPEP as an intervention, the only validation study involved a regression of initial SPEP ratings on recidivism rates over the following six and 12 months with an early version of SPEP (which included no rating for the quality factor) that found lower than predicted rates of recidivism for programs with higher SPEP scores (Lipsey, 2008; Redpath &

Brandner, 2010). At present, there is no implementation evidence that an improvement in SPEP scores is related to a reduction in recidivism rates.

4.1 | Impact on the juvenile justice system

The second criterion for determining the relative value of program and practice approaches (the first is the evidence base for programs and practices, discussed above) is to assess their impact on juvenile justice programming—the extent of uptake and integration into service systems for juvenile justice and their impact on population rates of recidivism. For practices, we focus on the impact of SPEP as it is the only form of practice to date that has been implemented with scaling-up intent and embedded in juvenile justice systems.

5 | PROGRAM APPROACH

As noted earlier, the evidence base for MST, FFT, and TFCO is strong. These programs have been adopted and implemented successfully in hundreds of sites (Greenwood, 2018). As of 2011, MST has been implemented in 34 states, 15 countries, and 540 sites worldwide, and reports serving 17,000 families annually (MST Services, 2019).⁴ Functional Family Therapy (FFT) has been implemented in 24 states and 270 sites, and treats 17,500 youth and families annually. Treatment Foster Care Oregon (TFCO) has been implemented in 13 states and 50 sites, and serves approximately 1,300 youth annually. Both MST and FFT have been implemented *statewide* in justice systems in eight states, and their implementation levels are high in at least five additional states (Colorado Office of State Planning and Budgeting, 2018; Elliott & Fagan, 2017; Greenwood, 2018; Henggeler & Schoenwald, 2011; Welsh & Greenwood, 2015). While the number of juvenile justice systems involved and the numbers of offenders served are significant and growing, on a national scale the current adoption level of evidence-based programs and the proportion of offenders served by these programs is small.

The best evidence for the impact of evidence-based programs on juvenile justice systems is the population-level reductions in recidivism and cost savings achieved in statewide implementations of evidence-based programs. The results achieved in Washington's 1997 Community Juvenile Justice Act was a 10% statewide reduction in recidivism, which will produce a projected state cost savings of \$425 million by 2030. Based largely on the success of this initiative, the Washington State legislature cancelled the planned construction of a new state prison (Barnoski, 2004; Drake et al., 2009). Project Redirection in Florida achieved a statewide reduction in total recidivism of 8% and a 24% reduction in felony recidivism, and a cost savings of \$124 million in 2010–2011. Pennsylvania's Research-Based Programs Initiative resulted in a \$317 million cost savings (Jones et al., 2008). These are significant population-level effects at the state level.

In addition, these evidence-based programs and the program approach in general have affected how states, philanthropic organizations, and countries are funding and administering juvenile justice programming (Greenwood, 2018). There is increasing evidence that states and local agencies are willing to scale up these programs. The success of the Washington State initiative led to a national initiative, *Results First* (<https://www.pewtrusts.org/en/projects/results-firstinitiative>), a joint project of the Pew Charitable Trust Foundation and the MacArthur Foundation, that provides states across the country with technical assistance to conduct similar reviews of their use

of evidence-based programs and their cost benefits. The programs selected for this initiative have been predominantly Blueprint model programs.

Not only are these evidence-based programs being adopted and implemented, they are central to current innovations in juvenile justice systems. Both MST and FFT are approved for Temporary Assistance for Needy Families (TANF) funding and have been rated as “well-supported” programs by the Title IV-E Family First Preventive Services Act, which provides federal funding to states that implement these two programs. Florida’s Project Redirection was awarded the Prudential Financial-Davis Productivity Award for innovations in case processing and the 2008 Science and Service Award from the Substance Abuse and Mental Health Services Administration for its success in helping communities adopt and implement effective programs (Elliott & Fagan, 2017). This initiative is now being implemented in Georgia with two Blueprint programs—FFT and a promising Blueprints program, The Strengthening Families 10–14 Program (a group parenting and youth skills program that aims to reduce aggressive, hostile behavior, and substance abuse in adolescents).

Unfortunately, we know of little reliable data on the adoption rates at the national level or effects on system-level recidivism for the other evidence-based programs for juvenile justice populations. Many states, however, are in the process of developing an inventory of programs being implemented in their juvenile justice systems (Greenwood, 2018), so these data may be available in the near future. Lipsey (2018) estimates that evidence-based programs in the Blueprints, OJP, and OJJDP registries constitute less than 10% of programs currently being implemented in the states where SPEP has been involved in creating local juvenile justice system inventories of programs. It is clear that evidence-based programs have not yet had a major national effect on juvenile justice programming, but it is equally clear that they have had a significant impact at the state level and demonstrate the potential for achieving a national-level impact.

6 | PRACTICE APPROACH: SPEP

Greenwood (2018) reports that as of 2017, SPEP was being implemented in 10 states. Three of these states were involved in the OJJDP Juvenile Justice Reform and Reinvestment Initiative (JJRRI) in a demonstration project, in which SPEP was implemented in: 1) a single county in Wisconsin, 2) three judicial districts in Iowa, and 3) a set of programs in Delaware serving high-risk offenders that had not previously been evaluated or audited by the state. The implementation of SPEP in Pennsylvania was statewide but we have no information about the level of implementation in the other states.

While this is a small number of sites and the level of implementation in the demonstration sites is limited, SPEP potentially can involve multiple juvenile justice systems, a number of different types of practices, and many separate programs, and it can serve a large number of offenders. However, we know of no reliable estimates of the total number of juvenile justice systems, programs, or individual offenders that have been impacted by SPEP. The number of states in which SPEP is being implemented is substantially smaller than that for the four evidence-based programs considered above, and to our knowledge SPEP has not been implemented outside the United States, but there is no current basis for comparing the juvenile justice system adoption rates of evidence-based programs versus practices or their impact on the number of offenders served.

To date, SPEP also has had a small level of adoption by juvenile justice systems. There is no experimental evidence that this practice-focused intervention has had a significant impact on

juvenile justice program effectiveness or overall system effectiveness in those sites where it has been implemented.

Overall, evidence-based interventions have had limited national impact. As discussed in the previous section, evidence-based programs have shown reduced recidivism rates in selected states, with a significant number of adoptions occurring both nationally and internationally. The impact of evidence-based practices (and SPEP specifically), however, is essentially unknown. It should be noted that in addition to the juvenile justice evidence-based programs (MST, FFT and TFCO), other programs are rated as model+ on the Blueprints registry and are serving large numbers of youth involved in delinquent behavior and those at risk for involvement: Life Skills Training (LST), a middle-school substance abuse prevention program; Positive Action (PA), a social emotional learning program for students in elementary and middle-schools; and Project Towards No Drug Abuse (TND), a high-school drug prevention program. For example, as of 2017, LST has been implemented in more than 1,200 communities serving more than a million youth with a cost-benefit ratio of \$25.61 for every \$1 spent (Center for the Study and Prevention of Violence, 2018; Jensen & Hawkins, 2018; Washington State Institute for Public Policy, 2017). The evidence indicates that evidence-based programs for reducing juvenile crime are being scaled up with fidelity and are having demonstrable impact. And there is reason to believe the adoption rate of these programs will continue to expand given their successful implementation and impact on recidivism (Jensen & Hawkins, 2018).

6.1 | Fidelity, effect size, and cost considerations

One argument offered for practices versus programs is that programs require fidelity to specific implementation that is too challenging for juvenile justice systems (Embry & Biglan, 2008; Lipsey, 2018; Ringwalt et al., 2003). This position asserts that practices can be more readily integrated into existing programming and levels of staff training. Two other considerations in judging how advantageous a given program or practice might be are the program's effect size and cost (and cost-benefit ratio). While there are many complex issues related to adoption decisions and successful implementation, we focus here on three issues—fidelity, effect size, and cost—and compare the evidence for evidence-based programs and practices (the single available example of practices being SPEP).

7 | FIDELITY: THE PROGRAM APPROACH

In general, some researchers claim that the implementation of evidence-based programs is more highly specific, leaves less room for local adaptation, and is less flexible than practice approaches such as SPEP (Barnoski, 2002; Fixen, Blasé, Metz, & Van Dyke, 2013; Greenwood, 2018; Lipsey, 2018). This raises questions about the difficulty of implementing these interventions with fidelity. Achieving and maintaining high levels of fidelity is a challenge in scaling up an intervention, particularly for programs because they are fully prescriptive (i.e., designed to be implemented as specified), whereas practices carry a presumption of being more flexible, allowing for some variation in delivery, dosage, and other non-core components and thus being more amenable to local adaptations.

Achieving fidelity is a ubiquitous problem in implementation science (Brownson, Colditz, & Proctor, 2018; Fixen, Naom, Blasé, Friedman, & Wallace, 2005; Spoth et al., 2013). How to attain

needed fidelity and how close that fidelity must be to realize a program's expected effectiveness are still unresolved issues. Yet, available research indicates that when programs are continually monitored for fidelity with a high level of developer involvement, as is the case for the four Blueprint juvenile justice model programs, scale-up fidelity can be maintained and positive program effectiveness achieved (Durlak & DuPre, 2008; Elliott & Mihalic, 2004; Fixen et al., 2005; Kemp, 2016; Mihalic, Fagan, & Argamaso, 2008; Rhoades, Bumbarger, & Moore, 2012; Welsh, Sullivan, & Olds, 2010). Both MST and FFT have demonstrated implementation fidelity and positive levels of effectiveness in scale up and routine practice as described earlier, and they have confirmed the expected relation between therapist and supervisor fidelity and recidivism (Kemp, 2016; Mihalic et al., 2002; MST Services, 2019; Ogden et al., 2012; Schoenwald, Sheidow, Letourneau, & Liao, 2003; Schoenwald, Chapman, Sheidow, & Carter, 2009; Washington State Institute for Public Policy, 2004).

In a national dissemination and replication study of Blueprint violence prevention model programs involving 42 sites, Elliott and Mihalic (2004) found 74% of these model programs implemented *all* the core components of their respective programs, and the average for all programs was 86% to 100% percent of core components. This study also included an implementation of LST in 105 sites involving approximately 430 middle schools. Implementation fidelity scores averaged between 81% and 86% of required lesson content. A more recent LST replication initiative involving 17 states, 724 schools, and more than 500,000 students found 68% to 87% of key lesson content was covered across all districts, and 92% of eligible students received the intervention (Center for the Study and Prevention of Violence, 2018). In all of these replication studies the monitoring of implementation fidelity was ongoing. All Blueprint model programs also have fidelity checklists available for monitoring during implementation.

8 | FIDELITY: THE PRACTICE APPROACH

Apart from SPEP, there has been no attempted scale-up of a juvenile justice practice and thus no assessment of implementation fidelity during scale-up. This issue has not been addressed to date; there has been no discussion of how to define or measure fidelity when implementing a practice.

What is the evidence that initial SPEP assessments of a practice's implementation fidelity improve over time with the use of SPEP? What is the evidence that when the SPEP-directed plan for improving fidelity is successfully completed, recidivism rates are reduced? To date no studies that provide evidence that the use of SPEP assessments to guide juvenile justice program upgrades results in increased program fidelity and effectiveness in any scale-up in the juvenile justice system. In fact, the research from the JJRRI SPEP demonstration program revealed several problems with the implementation of SPEP that suggest it will be difficult to implement SPEP assessments and plans for improving practice fidelity (Lieberman & Hussemann, 2016, 2017). For example, it took more than two years to complete the initial (baseline) rating of existing programs and install the systemwide data collection systems needed to assess improvement in the SPEP rating. Further, local agencies resisted the rigidity of the SPEP assessment requirements. Unfortunately, the demonstration grant ended before subsequent SPEP assessments could be completed. Implementing SPEP in this demonstration project was more difficult than expected. Achieving fidelity when scaling up SPEP may, therefore, prove to be as difficult as for evidence-based programs.

Research shows that fidelity of implementation can have a powerful impact on whether an intervention has intended or previously demonstrated benefits. Meeting the challenges of fidelity when implementing to scale, therefore, deserves serious attention. This includes recognizing that achieving high fidelity remains a challenge in most instances. And much still needs to be

established about the relation between fidelity and the expected effectiveness of an intervention. For example, some evidence suggests that this relation may involve a threshold effect rather than a simple linear effect—that is, some minimum level of fidelity required to achieve the expected effect. Berkel, Mauricio, Schoenfelder, and Sandler (2011), in a systematic review of studies evaluating the relation between implementation characteristics and outcomes, report that when implementation fidelity was at least 60%, programs reliably produced positive effects. Improving fidelity may not always result in a corresponding improvement in effectiveness as assumed in the SPEP intervention. Unfortunately, little is known about whether scaling up practices can be achieved with fidelity and with the expected levels of effectiveness. In contrast, there is clear evidence that some evidence-based programs have been scaled up with fidelity, achieve expected levels of effectiveness, and are cost-effective. There is no reliable evidence of effectiveness for some other evidence-based programs, however, and several have been found to be ineffective or iatrogenic.

8.1 | Effect size and cost-effectiveness

Decisions to adopt a program or practice involve more than considering just the extent and consistency of positive findings across multiple evaluations with scale-up and evidence that the intervention can be implemented with fidelity. Here we consider evidence of the effect size and cost savings of evidence-based programs and practices. Since we know of no data on scale-up of juvenile justice practices apart from SPEP and there is no evidence of effect size or cost-benefit for SPEP, this section is limited to evidence-based programs.

Earlier we reported on the effect sizes and cost-benefits of specific examples of the scale-up of MST, FFT, and TFCO. Based on a meta-analysis of all evaluation studies of each of these programs, both MST and FFT produced positive average effect sizes and cost-benefits. Washington State Institute for Public Policy (WSIPP) estimated the cost-benefit of FFT when implemented with youth in state institutions to be \$18.75 for every \$1 spent, with an adjusted effect size of $d = -.25$ (adjusted from empirical findings on effects when taken to scale). When FFT was implemented with youth on probation, the cost-benefit ratio was \$8.35, with a similar adjusted effect size. For both populations, the probability of obtaining positive effects when implementing FFT with fidelity is 100%. For MST, WSIPP estimated a positive cost-benefit ratio of \$3.02 and adjusted effect size of $d = -.10$, and a probability of a positive cost-benefit of 99% percent. The cost-benefit for TFCO was estimated to be \$4.29 with a probability of a positive effect being 91% (Washington State Institute for Public Policy, 2017, 2019). Cost savings can vary from state to state depending on local conditions. The estimated cost-benefit ratio for MST statewide in New Mexico is \$3.34 (Dopp et al., 2018), and in Colorado it is \$3.00 (Colorado Office of State Planning and Budgeting, 2018). These cost-benefit estimates include the initial start-up costs for evidence-based programs.

There is also evidence from the Washington State Institute for Public Policy (2019) on the cost-effectiveness of several other evidence-based programs. The Adolescent Diversion Program has a 100% probability of benefits exceeding costs; MST for Substance Abuse has a cost-benefit ratio of 1.58 with a 58% probability of a positive effect; and Aggression Replacement Training has a cost-benefit ratio of 1.51 with a 23% probability of benefits exceeding costs. WSIPP recommends a 75% probability as the standard for being a cost-effective intervention, suggesting that these latter three programs are not cost-effective.

As with fidelity, there is no study that identifies the cost-benefit of different practices primarily because such evaluations are lacking in the source studies. Nor has there been such an evaluation of the SPEP approach or for the practices identified by CrimeSolutions.

9 | DISCUSSION

9.1 | Scientific standard for evidence-based interventions

One of the problems with the evidence-based program approach is the variability in registry standards for certifying interventions as evidence-based (Fagan & Buchanan, 2016; Means, Magura, Burkhardt, Schröter, & Coryn, 2015). This variability is confusing and results in different levels of confidence that can be placed in the effectiveness of programs recommended for scale-up. The problems emanating from this variation are exacerbated when meta-analytic methods use differential methodological criteria for including studies than program registries use.⁵

The recommended standard for demonstrating the effectiveness of a prevention program established by the National Academy of Science (NAS) requires multiple well-conducted randomized controlled trials. Further, the NAS recommendation cautions that when this type of evidence is not available, evidence from the next strongest designs, i.e., quasi-experimental designs, should not be considered definitive (National Research Council and Institute of Medicine, 2009). Design standards that require any included evaluations to meet basic benchmarks for causal interpretation (RCTs) or be scrutinized for low likelihood of bias (QEDs with strong controls for potential bias) are needed to align registry ratings and facilitate understanding of programs and practices comparatively.

In addition to the design quality, replication, and independence of evaluations, the scientific standard for certifying an intervention as evidence-based should also require specification of the causal mechanism or process underlying the intervention. We agree with Nagin and Sampson (2019), Sampson, Winship, and Knight (2013), Weisburd et al. (2017), and Weisburd, Hinkle, Braga, and Wooditch (2015), who warn of the danger associated with basing policy and programming decisions or recommendations for scale-up on interventions that have no clear indication of the causal mechanism operating to produce the observed effect. Without this, evidence for a generic practice's average effect cannot direct juvenile justice programming to any meaningful extent.

The Blueprint registry standard is close to the NAS standard for certifying programs as evidence-based. For its model and model+ programs, Blueprints requires a logic model, at least one high-quality RCT, a replication of a high-quality RCT or QED, a one-year sustained effect, no iatrogenic effects, and an independent evaluation for a model+ rating. The two Department of Justice registries are further from the NAS standard, allowing for a program with a single high-quality QED to be certified as an evidence-based (i.e., effective) program. This is not a problem limited to the evidentiary standard for crime and delinquency interventions certified as evidence-based. It is a general problem for the field of intervention science. The inconsistency in certification standards is one of the factors that undermines practitioner confidence in the use of registries (Nuehoff et al., 2015). We need consensus on a minimum standard for certifying programs and practices as evidence-based and ready for scale-up. At present, there is too little empirical evidence about any given practice, let alone the approach overall, to undertake judgment about its potential value, whether in comparison to program approaches or simply as another viable approach.

10 | IMPACT ON JUVENILE JUSTICE SYSTEMS

The impact of evidence-based programs on juvenile justice programming is not yet what is hoped for. However, the rate of adoption and scale-up of these programs to date is quite like that of

prevention interventions in other fields of prevention science (Glasgow et al., 2012; Morris, Wooding, & Grant, 2011; Welsh et al., 2010). In reviewing program evaluation and implementation in education, health, justice, and other social policy areas, Shadish, Cook, and Leviton (1991) note the difficulty of introducing significant change in established institutional practice and that, as a result, policies or programs that call for more than marginal changes in the status quo are rarely implemented to scale. It is naive to assume that once a new innovative program or practice has been demonstrated to be more effective than existing institutional practice it will be widely disseminated within a year or two, and within a few more years will be firmly embedded in any institutional system, including the juvenile justice system. Historically, the dissemination and adoption of innovative programs and practices like those identified as evidence-based programs are relatively slow, not only in juvenile justice, but in health, education, and other institutional systems.

For example, it was more than 100 years after Ignaz Semmelweis documented the effectiveness of surgeons washing their hands to avoid infections that the first national hand hygiene guidelines were published (World Health Organization, 2009); it was 40 years after James Lind established that vitamin C was an effective preventive for scurvy before the British Admiralty ordered British ships to carry a supply of lemon juice on all voyages (http://www.bbc.co.uk/history/historic_figures/lind_james.shtml); and after it was discovered that mosquitos were carriers of malaria it took more than a decade before fumigation use as a deterrent to malaria became routine (McCullough, 1977). In describing change in social policy, Friedman and Friedman (1989) assert that the process leading to a change in the institutional status quo and the establishment of new policies takes decades. By this standard, we believe the level of juvenile justice system adoption of evidence-based programs is on track and with continued support and development, it will soon come to have a truly significant impact on the effectiveness of juvenile justice system interventions nationally.

11 | NUMBER OF EVIDENCE-BASED PROGRAMS AND PRACTICES

Another frequent criticism of the evidence-based program approach related to its limited impact is the small number of programs certified as evidence-based (Embry & Biglan, 2008; Lipsey, 2018). The appropriate question is *how many evidence-based programs does it take to have a significant impact on the juvenile justice system, and to have a significant impact on juvenile justice systemwide recidivism rates?* In the Washington State initiative described earlier, only four programs were involved. While the proportion of all offenders in Washington's juvenile justice system involved in these programs was not reported, it seems likely it was most offenders. Project Redirection in Florida included only three programs. The Research-Based Program Initiative in Pennsylvania included seven. Lipsey's (2009) meta-analysis of the full range of programs being implemented in juvenile justice systems identified four types of generic practices that were effective. Presumably, these four practice types captured the full range of services being delivered by juvenile justice systems prior to 2003. A substantial impact on the juvenile justice system does not depend on having a large number of programs. It does not depend on having programs with large effect sizes. It can be achieved with relatively few programs that meet a high evidentiary standard of effectiveness, can be implemented with fidelity, and are cost-effective. We have evidence-based programs that meet these requirements and, at present could, with selectivity in programs, successfully replicate the experience in Washington State throughout the United States.

Obviously, having more evidence-based programs and offering a broader range of services appropriate for a broader range of offender needs are still important goals. This would provide more choice to practitioners, increase the likelihood of a good fit between program characteristics and local needs, and expand the numbers of offenders in the juvenile justice system being served by effective interventions.

12 | STANDARDS FOR CERTIFYING EFFECTIVE PRACTICES

We are far from identifying core components or the specific intervention elements that account for the observed changes in targeted outcomes (MacKenzie & Farrington, 2015; Weisburd et al., 2015). This is true for evidence-based programs as well as for practices. Not only would good research evidence on core components facilitate the development of evidence-based practices, it would validate specific theories and effectiveness of specific change strategies employed to modify risk and protective factors. It would also inform local adaptations to programs to insure they are consistent with the causal mechanism involved so as not to undermine the effectiveness of the intervention. We support and encourage the adoption of the Institute of Medicine's (2011) recommended Standards for Systematic Reviews that specifically addresses this issue, requiring that the selection of the topic for the review involve an analytic framework that clearly identifies the chain of logic that links the intervention to the outcome of interest (Standard 2.5).

In addition, a set of evidentiary standards is needed for certifying practices as evidence-based. This will require more specificity in selection requirements for meta-analyses attempting to identify generic types of programs, with attention to common core components, intervention delivery strategies, and theorized causal mechanisms. There are examples of meta-analyses that selected adult justice system programs based on common core components theoretically linked to the outcome—for example, Pulling Levers (Braga & Weisburd, 2012) and Job Transitions programs (Barden, Juras, Redcross, Farrell, & Bloom, 2018)—but the use of this selection criterion is rare and we know of none involving juvenile justice practices. And, while these examples define inclusion based on some presumed shared causal logic, the value of such approaches still is hindered by inclusion of studies with insufficient methodological quality to instill confidence about the results. An unresolved issue is how a practice that includes interventions with high quality evaluations demonstrating significant negative effects should be considered in meta-analyses and in the certification of a practice as evidence-based. The existing registries agree that a program with a well-conducted experimental trial finding a negative effect is not certified as an effective or model program even if the average effect from the set of trials is positive. Developing standards for identifying evidence-based practices should be a high priority.

13 | POLICY IMPLICATIONS

The contention of some that the evidence-based program approach has failed to have substantial impact on juvenile justice is valid if the standard is that evidence-based programs serve most offenders in the juvenile justice system. Though that appears to be the case in Washington State, this level of broad impact has not yet been demonstrated consistently. On the other hand, the contention that reliance on practices rather than evidence-based programs overcomes the forces limiting program use and impact does not have any solid examples or convincing evidence.

Considering the existing evidence, we make some suggestions below about how to advance the impact of evidence-based interventions, programs, and practices in juvenile justice.

At present, we think funding for the scale-up of interventions to upgrade the effectiveness of juvenile justice programming should be limited to evidence-based programs and, more specifically, to those that meet a high standard requiring well-designed and implemented RCTs showing sustained effects and replication. The evidence for practices as effective interventions in the juvenile justice system is too limited for any confident judgement about their utility. At this point, there is no sound scientifically rendered support for any specific practices, including SPEP, to be considered ready for scale-up. More attention to how programs are collectively analyzed to identify practices and experimental tests of their effectiveness in scale-up is needed before such judgement can be rendered.

This is not to suggest that only evidence-based programs should be promoted. Current research on practices provides an important body of research that can build theory, guide program development, and inform local adoption decisions when there are no evidence-based programs available that fit local needs. But such decisions must be made with full awareness that the current evidence for these interventions is substantially lower and the risk of null effects is substantially greater than for evidence-based programs. Moreover, this option should not be considered when planning for a major scale-up. Interventions with a lower level of evidence, like SPEP, should be designated and promoted as promising or as a research-based intervention (Washington State Institute for Public Policy, 2017).

14 | CONCLUSION

Fixen et al. summarize the current state of intervention science as follows: “Attempts to scale ineffective or harmful innovations are a waste of time, money and opportunity. Yet investing in strategies that do not work is the norm and not the exception as well-meaning legislators and leaders press for quick solutions in human service systems” (2017, p. 494). If we are to change this situation, we must be able to identify and implement interventions that are *known to work, that can be implemented at scale with fidelity and that are cost-effective*. The evidence-based program approach has provided this kind of evidence for selected programs and demonstrated success in scaling them up with significant impact on juvenile justice systems.

The generic practice approach to prevention is currently underdeveloped and has not qualified any practices as evidence-based. Conceptual and construction problems must be addressed; the current research evidence for practices is weak. There is no experimental evidence to date that this approach has identified practices that can be scaled-up to impact juvenile justice programming.

Relatively few evidence-based programs for adjudicated juvenile offenders currently exist, but some have been shown to work, are cost-effective, and can be implemented at scale with fidelity. While the number of these evidence-based programs is small, these programs’ impact on the few juvenile justice systems implementing them has been substantial, and the number of systems adopting these evidence-based programs is accelerating.

Given the history of the time it takes for effective interventions to become embedded in institutional routine practice, the impact of these interventions on juvenile justice programming is on track. What holds the greatest promise for improving the effectiveness of the juvenile justice system is more high-quality research on promising and research-based programs and practices that will provide the additional evidence needed to qualify them as evidence-based interventions.

CONFLICT OF INTEREST

Believing in the need for full disclosure in Prevention Science, the authors certify that they are all members of the Blueprints Advisory Board and receive compensation for time spent on program reviews and Board meetings; that they have no financial or other conflict of interest with respect to any of the specific interventions discussed in this article.

ORCID

Delbert S. Elliott  <https://orcid.org/0000-0001-7386-8896>

Pamela R. Buckley  <https://orcid.org/0000-0002-8268-3524>

Denise C. Gottfredson  <https://orcid.org/0000-0003-3075-8392>

Patrick H. Tolan  <https://orcid.org/0000-0001-5669-8442>

ENDNOTES

- ¹ In reviewing the evidence for *Treatment in Secure Corrections for Serious Juvenile Offenders*, we found it difficult to ascertain the common features that characterize it and how fidelity to the practice might be measured. The practice is described as including clinical, psychological, educational and behavioral programs.
- ² In describing what “effective implementation” of a program means, Lipsey notes both ensuring the program has the distinguishing characteristics of similar programs that have above average positive effects (the four characteristics identified) and *avoiding those characteristics found to have negligible or negative effects*. But there is no indication these latter characteristics were investigated, or any indication in the therapeutic types or SPEP assessment that these characteristics were considered or used to disqualify a therapeutic type program from using SPEP as a way of improving its fidelity and claiming improved effectiveness (Lipsey et al., 2010).
- ³ Personal correspondence with Mark Lipsey: email May 28, 2016.
- ⁴ MST claims 200,000 families have been positively impacted by MST (mstservices.com).
- ⁵ Gottfredson et al. (2015) have proposed specific standards for efficacy, effectiveness, and scale-up in prevention science that could and should guide program and practice identification for evidence-based approaches in criminal justice.

REFERENCES

- Barden, B. R., Juras, R., Redcross, C., Farrell, M., & Bloom, D. (2018). *New perspectives on creating jobs: Final impacts of the next generation of subsidized employment programs* (The Enhanced Transitional Jobs Demonstration). MDRC.
- Barnoski, R. (2002). *Washington State's implementation of functional family therapy for juvenile offenders: Preliminary findings* (Document No. 02-08-1201). Washington State Institute for Public Policy.
- Barnoski, R. (2004). *Outcome evaluation of Washington State's research-based programs for juvenile offenders* (Document Number: 04-01-1201). Washington State Institute for Public Policy.
- Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science, 12*(1), 23–33.
- Braga, A. A., & Weisburd, D. L. (2012). The effects of “pulling levers” focused deterrence strategies on crime. *Campbell Systematic Reviews, 8*(1), 1–90.
- Brownson, R. C., Colditz, G. A., & Proctor, E. K. (Eds.). (2018). *Dissemination and implementation research in health: Translating science to practice* (2nd ed). Oxford University Press.
- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Center for the Study and Prevention of Violence. (2018). All sites year-end report 2017–2018. *Blueprints for Healthy Youth Development: LifeSkills Training Program Replication Project*. Boulder, CO: University of Colorado.
- Chaplin, D. D., Cook, T. D., Zurovac, J., Coopersmith, J. S., Finucane, M. M., Vollmer, L. N., & Morris, R. E. (2018). The internal and external validity of the regression discontinuity design: A meta-analysis of 15 within-study comparisons. *Journal of Policy Analysis and Management, 37*(2), 403–429.

- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283–292.
- Colorado Office of State Planning and Budgeting (2018). *Colorado Multi-Systemic Therapy (MST) pilot project in underserved regions*. Denver, CO: Colorado Office of State Planning and Budget.
- 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.
- Coyne, J., Thombs, B. D., & Hagedoorn, M. (2010). Ain't necessarily so: Review and critique of recent meta-analysis of behavioral medicine interventions in health psychology. *Health Psychology*, 29(2), 107–116.
- Dopp, A. R., Coen, A. S., Smith, A. B., Reno, J., Bernstein, D. H., Kerns, S. E. U., & Altschul, D. (2018). Economic impact of the statewide implementation of an evidence-based treatment: Multisystemic therapy in New Mexico. *Behavior Therapy*, 49, 551–566.
- Drake, E. K., Aos, S., & Miller, M. G. (2009). Evidence-based public policy options to reduce crime and criminal justice costs: Implications in Washington State. *Victims and Offenders*, 4, 170–196.
- Durlak, J., & DuPre, E. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3–4), 327–350.
- Durlak, J. A., & Lipsey, M. W. (1991). A practitioners guide to meta-analysis. *The American Journal of Community Psychology*, 19, 291–332.
- Elliott, D. S. (2017). Youth in forensic services: Evidence-based treatment of violence and aggression. In P. Sturme (Ed.), *The Wiley handbook of violence and aggression* (pp. 1–17). New York: John Wiley and Sons.
- Elliott, D. S., & Fagan, A. (2017). *The prevention of crime*. New York: Wiley-Blackwell.
- Elliott, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science*, 5, 47–53.
- Embry, D. D., & Biglan, A. (2008). Evidence-based kernels: Fundamental units of behavioral influence. *Clinical Child and Family Psychology Review*, 11(3), 75–113.
- EPISCenter. (2014). *Multisystemic Therapy in Pennsylvania: Three years of data, fiscal years 2012–2014*. Pennsylvania Commission on Crime and Delinquency, Pennsylvania Department of Human Services, PennState. Retrieved from <http://www.episcenter.psu.edu/sites/default/files/ebp/MST-Three-Year-Report-FINAL.pdf>
- EPISCenter. (2015). *Functional Family Therapy in Pennsylvania: Three years of data, fiscal years 2012–2014*. Pennsylvania Commission on Crime and Delinquency, Pennsylvania Department of Human Services, PennState. Retrieved from <http://www.episcenter.psu.edu/sites/default/files/ebp/FFT-Three-Year-Report-FINAL.pdf>
- Fagan, A. A., & Buchanan, M. (2016). What works in crime prevention? Comparison and critical review of three crime prevention registries. *Criminology & Public Policy*, 15(3), 617–649.
- Fixen, D., Blasé, K., Metz, A., & Van Dyke, M. (2013). Statewide implementation of evidence-based programs. *Exceptional Children*, 79(2), 213–230.
- Fixen, D. L., Blasé, K. A., & Fixen, A. A. M. (2017). Scaling effective innovations. *Criminology and Public Policy*, 16(2), 487–499.
- Fixen, D. L., Naom, S. F., Blasé, K. A., Friedman, R. M., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231).
- Friedman, M., & Friedman, R. D. (1989). The tide in the affairs of men. In A. Anderson & D. L. Bark (Eds.), *Thinking about America: The United States in the 1990's*. Stanford, CA: The Hoover Institution.
- Glasgow, R. E., Vinson, C., Chambers, D., Khoury, M. J., Kaplan, R. M., & Hunter, C. (2012). National Institutes of Health approaches to dissemination and implementation science: Current and future directions. *American Journal of Public Health*, 102(7), 1274–1281.
- Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention science*, 16(7), 893–926.
- Gorman, D. M., & Conde, E. Huber Jr, J. C. (2007). The creation of evidence in 'evidence-based' drug prevention: A critique of the Strengthening Families Program Plus Life Skills Training Evaluation. *Drug and Alcohol Review*, 26(6), 585–593.

- Greenwood, P. W. (2018). Ranking states in their use of evidence-based programs for juvenile offenders: A 20-year progress report. *Berkeley Journal of Criminal Law, 23*, 242–260.
- Henggeler, S. W., & Schoenwald, S. K. (2011). Evidence-based interventions for juvenile offenders and juvenile justice policies that support them. *Social Policy Report, 25*(1), 1–28.
- Institute of Medicine. (2011). *Finding what works in health care: Standards for systematic reviews*. Washington, DC: National Academies Press.
- Jaffee, S. R., Strait, L. B., & Odgers, C. L. (2012). From correlates to causes: Can quasi-experimental studies and statistical innovations bring us closer to identifying the causes of antisocial behavior? *Psychological Bulletin, 138*(2), 272–295.
- Jensen, J., & Hawkins, J. D. (2018, May 1). *Seven ways to unleash the power of prevention* [Presentation]. Blueprints for Healthy Youth Development Annual Conference, Westminster, CO.
- 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*. Prevention Research Center for the Promotion of Human Development, Pennsylvania State University.
- The Justice Research Center. (2013). *Redirection: Consistently reducing juvenile crime in Florida*. Tallahassee, FL: The Justice Research Center.
- Kemp, L. (2016). Adaptation and fidelity: A recipe analogy for achieving both in population scale implementation. *Prevention Science, 17*(4), 429–438.
- Knott, L., Wanner, P., & He, L. (2019). *Washington State's Aggression Replacement Training for juvenile court youth: Outcome evaluation*. (Document Number 19-06-1201). Olympia: Washington State Institute for Public Policy.
- Kraemer, H. C., Gardner, C., Brooks, III, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods, 3*(1), 23–31.
- Lieberman, A., & Huessmann, J. (2016). *Implementing the Standardized Program Evaluation Protocol™ to rate juvenile justice programs* (Research Report). Washington DC: The Urban Institute.
- Lieberman, A., & Hussemann, J. (2017). *Local validation of SPEP™ ratings of juvenile justice program effectiveness* (Research Report). Washington DC: The Urban Institute.
- Lipsey, M. W. (1988). Reports on practice areas: Practice and malpractice in evaluation research. *Evaluation Practice, 9*(4), 5–24.
- Lipsey, M. W. (2008). *The Arizona Standardized Program Evaluation Protocol (SPEP) for assessing the effectiveness of programs for juvenile probationers: SPEP ratings and relative recidivism reduction for the initial SPEP sample. A report to the Juvenile Justice Services Division, Administrative Office of the Courts, State of Arizona*. Nashville, TN: Vanderbilt Institute for Public Policy Studies, Vanderbilt University.
- Lipsey, M. W. (2009). The primary characteristics that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders, 4*(2), 124–147.
- Lipsey, M. W. (2014, May). *Evidence-based programs and practices: Examples from juvenile justice* [Presentation]. Society for Prevention Research Symposium. Programs, Practices and Beyond: Guiding Implementation Scientifically. Washington, DC.
- Lipsey, M. W. (2018). Effective use of the large body of research on the effectiveness of programs for juvenile offenders and the failure of the model programs approach. *Criminology and Public Policy, 17*(1), 189–198.
- Lipsey, M. W., Howell, J. C., Kelly, M. R., Chapman, G., & Carver, D. (2010). *Improving the effectiveness of juvenile justice programs*. Washington DC: Center for Juvenile Justice Reform.
- MacKenzie, D. L., & Farrington, D. P. (2015). Preventing future offending of delinquents and offenders: What have we learned from experiments and meta-analyses? *Journal of Experimental Criminology, 11*(4), 565–595.
- McCullough, D. (1977). *The path between the seas*. New York: Simeon and Schuster.
- Means, S. N., Magura, S., Burkhardt, J. T., Schröter, D. C., & Coryn, C. L. (2015). Comparing rating paradigms for evidence-based program registers in behavioral health: Evidentiary criteria and implications for assessing programs. *Evaluation and Program Planning, 48*, 100–116.
- Mihalic, S., Ballard, D., Michalski, A., Tortorice, J., Cunningham, L., & Argamaso, S. (2002). *Blueprints for Violence Prevention Violence Initiative: Final process evaluation report*. Center for the Study and Prevention of Violence, Institute of Behavioral Science, University of Colorado, Boulder.
- Mihalic, S. F., Fagan, A. A., & Argamaso, S. (2008). Implementing the LifeSkills Training drug prevention program: Factors related to implementation fidelity. *Implementation Science, 3*(1), 5.

- Morris, Z. S., Wooding, S., & Grant, J. (2011). The answer is 17 years, what is the question: Understanding time lags in translational research. *Journal of the Royal Society of Medicine*, *104*(12), 510–520.
- Mullen, E. J., & Streiner, D. L. (2004). The evidence for and against evidence-based practice. *Brief Treatment and Crisis Intervention*, *4*(2), 111–121.
- MST Services. (2019). *Research at a glance*. January 2019. Retrieved from www.mstservices.com/mst-whitepapers
- Nagin, D. S., & Sampson, R. J. (2019). The real gold standard: Measuring counterfactual worlds that matter most to social science and policy. *Annual Review of Criminology*, *2*, 123–145.
- National Research Council and Institute of Medicine. (2009). *Preventing mental, emotional, and behavioral disorders among young people: Progress and possibilities*. Washington DC: The National Academy Press.
- Neuhoff, A., Axworthy, S., Glazer, S., & Berfond, D. (2015). *The What Works marketplace: Helping leaders use evidence to make smarter choices*. Boston, MA: Bridgespan.
- Ogden, T., Bjørnebekk, G., Kjøbli, J., Patras, J., Christiansen, T., Taraldsen, K., & Tollefsen, N. (2012). Measurement of implementation components ten years after a nationwide introduction of empirically supported programs—a pilot study. *Implementation Science*, *7*(1), 49.
- Redpath, D. P., & Brandner, J. K. (2010). *The Arizona Standardized Program Evaluation Protocol (SPEP) for assessing the effectiveness of programs for juvenile probationers: Rating and relative recidivism reduction: An update to the January 2008 report by Dr. Mark Lipsey*. Phoenix, AZ: Arizona Supreme Court Administrative Office of the Courts, Juvenile Justice Service Division. Retrieved from my.vanderbilt.edu/spep/files/2013/04/AZ-SPEP-Recidivism-Analysis_Update_-2010.pdf
- Rehuher, D., Hiramatsu, T., & Helm, S. (2008). Evidence-based youth drug prevention: A critique with implications for the practice-based contextually relevant prevention in Hawai'i. *Hawai'i Journal of Public Health*, *1*(1), 52–61.
- Rhoades, B. L., Bumbarger, B. K., & Moore, J. E. (2012). The role of a state-level prevention support system in promoting high-quality implementation and sustainability of evidence-based programs. *American Journal of Community Psychology*, *50*, 386–401.
- Ringwalt, C. L., Ennet, S., Johnson, R., Rohrbach, L., Simons-Rudolph, A., Vincus, A., & Thorne, J. (2003). Factors associated with fidelity to substance use prevention curriculum guides in the nation's middle schools. *Health Education & Behavior*, *30*, 375–391.
- Rosenbaum, P. R., & Rubin, D. B. (1984). Reducing bias in observational studies using subclassification on the propensity score. *Journal of the American Statistical Association*, *79*, 516–524.
- Sampson, R. J., Winship, C., & Knight, C. (2013). Translating causal claims: Principles and strategies for policy-relevant criminology. *Crime and Public Policy*, *12*, 1–30.
- Schoenwald, S. K., Chapman, J. E., Sheidow, A. J., & Carter, R. E. (2009). Long-term youth criminal outcomes in MST transport: The impact of therapist adherence and organizational climate and structure. *Journal of Clinical Child and Adolescent Psychology*, *38*, 91–105.
- Schoenwald, S. K., Sheidow, A. J., Letourneau, E. J., & Liao, J. G. (2003). Transportability of Multisystemic Therapy: Evidence for multi-level influences. *Mental Health Services Research*, *5*, 223–239.
- Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991). *Foundations of program evaluation*. Newbury Park, CA: Sage Publications.
- Spoth, R., Rohrbach, L. A., Greenberg, M., Leaf, P., Brown, C. H., Fagan, A., . . . & Hawkins, J. D. (2013). Addressing core challenges for the next generation of type 2 translation research and systems: The translation science to population impact (TSci Impact) framework. *Prevention Science*, *14*(4), 319–351.
- Springer, J. F., & Phillips, J. (2007). The Institute of Medicine Framework and its implication for the advancement of prevention policy, programs and practice. *California, Community Prevention Initiative*, Center for Applied Research Solutions.
- Substance Abuse and Mental Health Services Administration. (2018). *Selecting best-fit programs and practices: Guidance for substance misuse prevention practitioners*. Retrieved from https://www.samhsa.gov/sites/default/files/ebp_prevention_guidance_document_241.pdf
- Walker, S. C., Lyon, A. R., Aos, S., & Trupin, E. W. (2017). The consistencies and vagaries of the Washington State Inventory of evidence-based practice: The definition of “evidence-based” in a policy context. *Administration and Policy in Mental Health and Mental Health Services Research*, *44*(1), 42–54.
- Washington State Institute for Public Policy. (2004, January). *Outcome evaluation of Washington State's research-based programs for juvenile offenders* (Document No 04-01-1201). Olympia, WA: WSIPP.

- Washington State Institute for Public Policy. (2017). *Updated inventory of evidence-based, research-based, and promising practices*. Olympia, WA: Washington State Institute of Public Policy. Retrieved from www.wsipp.wa.gov
- Washington State Institute for Public Policy. (2019, March). *Updated inventory of evidence-based, research-based, and promising practices for prevention and intervention services for children and juveniles in the child welfare, juvenile justice, and mental health systems*. Retrieved from <http://www.wsipp.wa.gov/BenefitCost>
- Weisburd, D., Farrington, D. P., Gill, C., Ajzenstadt, M., Bennett, T., Bowers, K., . . . Wooditch, A. (2017). What works in crime prevention and rehabilitation: An assessment of systematic reviews. *Criminology and Public Policy*, 16, 415–450.
- Weisburd, D., Hinkle, J. C., Braga, A. A., & Wooditch, A. (2015). Understanding the mechanisms underlying broken windows policing: The need for evaluation evidence. *Journal of Research in Crime and Delinquency*, 52, 589–608.
- Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science*, 578, 50–70.
- Welsh, B. C., & Greenwood, P. W. (2015). Making it happen: State progress in implementing evidence-based programs for delinquent youth. *Youth Violence and Juvenile Justice*, 13(3), 243–257.
- Welsh, B. C., Peel, M. E., Farrington, D. P., Elffers, H., & Braga, A. A. (2011). Research design influence on study outcomes in crime and justice: A partial replication with public area surveillance. *Journal of Experimental Criminology*, 7(2), 183–198.
- Welsh, B. C., Rocque, M., & Greenwood, P. W. (2014). Translating research into evidence-based practice in juvenile justice: Brand-name programs, meta-analysis, and key issues. *Journal of Experimental Criminology*, 10(2), 207–225.
- Welsh, B. C., Sullivan, C. J., & Olds, D. L. (2010). When early crime prevention goes to scale: A new look at the evidence. *Prevention Science*, 11, 115–125.
- World Health Organization. (2009). *WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge Clean Care is Safer Care*. Geneva: World Health Organization.

AUTHOR BIOGRAPHIES

Delbert S. Elliott is a Distinguished Professor Emeritus of Sociology, Fellow of the Institute of Behavioral Science, and Founder of the Center for the Study and Prevention of Violence at the University of Colorado Boulder. He is also the Founder and a current Advisory Board Member of Blueprints for Healthy Youth Development. His research interests include adolescent problem behavior and the prevention of violence, crime, and juvenile delinquency.

Pamela R. Buckley is a Research Associate and Fellow in the Institute of Behavioral Science at the University of Colorado Boulder and Director and Co-Principal Investigator of Blueprints for Healthy Youth Development. Her research interests include statistics, intervention research, evaluation methodology, research synthesis, and policy analysis. She has investigated and led trials of interventions designed to improve adolescent literacy skills and postsecondary readiness.

Denise C. Gottfredson is a Professor at the University of Maryland Department of Criminal Justice and Criminology and an Advisory Board Member of Blueprints for Healthy Youth Development. Her research interests include delinquency and delinquency prevention, and particularly the effects of school environments on youth behavior.

J. David Hawkins is Emeritus Endowed Professor of Prevention and Founder of the Social Development Research Group, School of Social Work, University of Washington. He is a Fellow

of the American Society of Criminology, the Society for Prevention Research, the American Academy of Social Work and Social Welfare, the Academy of Experimental Criminology, and the Washington State Academy of Sciences. In addition, he is an Advisory Board Member of Blueprints for Healthy Youth Development. He develops and tests preventive interventions for child and adolescent behavioral health problems involving families, schools, and communities as well as strategies for scaling tested and effective preventive interventions.

Patrick H. Tolan is Charles S. Robb Professor, Curry School of Education and Human Development and Professor, Department of Psychiatry and Neurobehavioral Sciences at the University of Virginia, where he is affiliated with the Youth-Nex UVA Center to Promote Effective Youth Development. He is also Emeritus Professor, University of Illinois and serves as an Advisory Board Member of Blueprints for Healthy Youth Development. He has led multiple developmental longitudinal and randomized control studies focused on prevention of academic, behavioral, and social failures and promoting resilience and effective youth functioning, with most emphasis on high-risk communities and/or critical developmental transitions.

How to cite this article: Elliott DS, Buckley PR, Gottfredson DC, Hawkins JD, Tolan PH. Evidence-based juvenile justice programs and practices: A critical review. *Criminol Public Policy*. 2020;19:1305–1328. <https://doi.org/10.1111/1745-9133.12520>