

❏ RESEARCH-BASED GUIDELINES FOR JUVENILE JUSTICE PROGRAMS

James C. Howell
The Comprehensive Strategy Group

Mark W. Lipsey
Peabody Research Institute
Vanderbilt University

❏ Abstract

Juvenile justice systems make use of many programs intended to reduce the recidivism of the juvenile offenders with whom they interact. Not all such programs are effective and one of the more progressive reforms of recent years has been the movement toward programs validated by research evidence. Three ways to define evidence-based programs are described, with a focus on a relatively unfamiliar approach—evidence from meta-analysis of evaluation research that supports the effectiveness of many generic types of programs. In contrast to the prevailing model program approach, this approach makes use of evidence that supports the effectiveness of many of the homegrown and local programs that juvenile justice systems use. The findings of a large meta-analysis of hundreds of studies reveal that many of these more generic programs are as effective as comparable model programs. These findings have been operationalized into a rating scheme based on the characteristics of effective interventions that can be used by service providers and juvenile justice systems to assess their programs. Two recidivism studies provide promising indications of the validity of this scheme for identifying effective programs and guiding improvement for ineffective ones. The results of this work show that the large body of research on interventions with juvenile offenders can be used to create guidelines that extend the concept of evidence-based programs to the kinds of generic programs most commonly used in juvenile justice systems.

The meta-analysis on which this paper is based was supported in part by grants from the National Institute of Mental Health, the Office of Juvenile Justice and Delinquency Prevention, and the Russell Sage Foundation. Thanks to the Arizona Juvenile Justice Services Division for their support of the recidivism analysis.

Efforts to implement evidence-based programs as a way to obtain better outcomes from juvenile justice interventions is arguably the most progressive policy reform of recent years. Not all programs are effective for producing the intended outcomes, but a program for which there is already evidence of effectiveness is generally a better bet than an untested one. Evidence-based programs can be defined in different ways, however, depending on what we mean by *program* and what we mean by *evidence*. Recognizing different definitions that nonetheless involve programs that are practical to implement and evidence that is scientifically credible can broaden the options for a juvenile justice system in useful ways. For this purpose, we distinguish three ways to view programs and their supporting evidence.

1. *The specific operating procedure of a particular program.*

A single, unique program is defined by its *specific operating procedure*—the implicit or explicit procedures that constitute that specific program in its particular site and context as practiced. Supporting evidence for any such specific program can be obtained by conducting an impact evaluation, that is, an evaluation that determines whether the program produces the intended outcomes. To provide the most valid results, an impact evaluation must use a control group of comparable juveniles who do not receive the program, preferably assigned randomly to program and no-program conditions. An evaluation of this sort for a specific program can provide credible evidence of effectiveness and, with positive results, that program can rightly claim to be evidence-based. Such direct evaluation, in fact, provides the most convincing evidence for a specific program. The disadvantage is the difficulty and expense of conducting such an evaluation, especially for the many programs used in a juvenile justice system.

2. *Brand name protocol programs.*

Another way of defining a program is by way of a manual or protocol that specifies how the program is to be implemented. There are many familiar examples of protocol programs in juvenile justice, e.g., Functional Family Therapy (FFT), Multisystemic Therapy (MST), Multidimensional Treatment Foster Care (MTFC), Aggression Replacement Training (ART), and the like. Using such a program requires that it be implemented locally with fidelity to the program developer's specifications for how it is to be delivered. Such programs are properly viewed as evidence-based if impact evaluations conducted on implementations elsewhere have found positive effects. These "model" or "exemplary" programs are typically identified through a review of research by some set of designated reviewers. Examples of such efforts include the Blueprints for Violence Prevention, the National Registry of Evidence-based Programs and Practices (NREPP), and the Office of Juvenile Justice and Delinquency Prevention (OJJDP) Model Programs Guide. Though their criteria vary and there is no consensus on the appropriate standards, programs appearing on lists such as these have become the *de facto* definition of what constitutes an evidence-based program. The advantages of this type of evidence-based program are the assurances of effectiveness if implemented with fidelity and the availability

of the program specifications, and, often, training and support from the developer to help with that implementation. Among the disadvantages is the possibility that the limited number of certified programs may not match all the needs of a juvenile justice system, the costs associated with purchasing these programs and the required support, possible resistance by established providers to changing to a new program, and uncertainty about whether the program will have the same effects in the local context as it had in the research sites.

3. *Generic intervention types.*

Programs can also be defined broadly according to the nature and focus of the services provided. Familiar examples include cognitive-behavioral therapy, family counseling, mentoring, victim-offender mediation, and the like. Within any such program *type*, specific programs vary somewhat but nonetheless meet the general criteria that make them more similar within a type than between types. For many such program types, there are multiple evaluations of individual programs that provide evidence about the effectiveness of programs of that general type. The most systematic approach to integrating the results of those studies is via a technique called meta-analysis. Meta-analysis characterizes the average effects found across the multiple studies and can identify some of the factors that differentiate the programs that produce the most positive effects from those less effective. A particular program can be called evidence-based if it is of a type for which multiple studies show positive average effects and it is implemented in a way that matches what those studies show to be the most effective versions of that program type. The advantage of this view of evidence-based programs is that it may include many established local programs that already are of a type supported by research or can be readily modified to match the research findings. The main disadvantage is that the research guidance for optimizing the effectiveness of different program types is itself rather generic. It lacks the specificity that comes with the protocol for a model program and the associated training and support systems that are also often available from the developer.

This article focuses on the last of these three approaches to evidence-based programming. The substantial body of evidence supporting the effectiveness of many generic interventions with juvenile offenders has not been well recognized, nor has its implications for evidence-based practice been fully appreciated. For many types of programs, this research shows that well-implemented generic versions perform as well as the brand name model programs that appear on the evidence-based practice lists. The challenge for making use of this large body of evidence for supporting effective programs in juvenile justice is to translate it into a form that provides sufficiently specific guidance about which types of programs are generally effective and how they should be implemented to attain the most positive effects.

With this in mind, Mark Lipsey and his colleagues have drawn on their large meta-analysis of research on the effectiveness of intervention programs for juvenile

offenders (Lipsey, 2009) to create the Standardized Program Evaluation Protocol (or SPEP for short). The SPEP is an evidence-based rating scheme for assessing the expected effectiveness of programs for reducing the recidivism of juvenile offenders. The SPEP creates a metric by assigning points to programs according to how closely their characteristics match those associated with the best recidivism outcomes for similar programs as identified in the meta-analysis (summarized below). In addition, the SPEP provides guidance with respect to program areas that need improving in order to achieve larger recidivism reductions and potentially greater effectiveness of a juvenile justice system's entire program continuum. We turn now to an overview of the findings of that meta-analysis, with particular attention to the exploration of the factors associated with variation in programs' effects on recidivism.

☒ Meta-Analysis of the Recidivism Effects of Interventions with Juvenile Offenders

The comprehensive meta-analysis that undergirds development of the SPEP best practice guidelines for juvenile delinquency programs was an effort begun by Lipsey in the mid 1980s and continues, with periodic updates, to the present day. The database of studies of interventions for juvenile offenders on which the meta-analysis was based includes information on 548 independent study samples drawn from 361 primary research reports. These studies represented all the intervention research that could be located through an extensive search for published and unpublished reports of research that met the following key criteria:

- The study involved juveniles aged 12–21 who received an intervention intended to have positive effects on their subsequent delinquency;
- At least one delinquency outcome was measured and assignment to treatment conditions was random or, if not, pretreatment differences were reported or matched.
- The study was conducted in an English-speaking country between 1958 and 2002.

Here we will only provide a summary of the main findings relevant to evidence-based guidance about effective interventions for adequately documented program types (further details can be found in Lipsey, 1992, 1999a, 1999b, 2009; Lipsey & Wilson, 1998). The analysis focused on identification of the program characteristics associated with the largest positive effects on recidivism. A central program characteristic in this regard was the generic type of intervention the program represented. Other categories of variables that were examined in relation to recidivism effects included:

- Characteristics of the juvenile samples, e.g., mean age, gender mix, predominant ethnicity, and delinquency risk.

- Juvenile justice supervision and control categorized as no supervision (prevention programs), diversion, probation or parole, and incarceration.
- Amount of service described in terms of the approximate duration of service (weeks from start to end) and the approximate number of total contact hours.
- Higher quality implementation, as indexed by involvement of the researcher in the monitoring and quality control of the service delivery and no mention of implementation problems in the study report, was associated with larger recidivism reductions. The involvement of the researcher was taken as a proxy for the extent to which attention was given to specifying the intended intervention, implementing it as intended, and monitoring implementation—all of which were mentioned as aspects of implementation quality in some study reports.
- Characteristics of the study methods for use as control variables to adjust for methodological differences among studies that might influence effect sizes. These included the way recidivism was measured, research design (randomized, matched, etc.), initial differences between intervention and control group, attrition, and information related to potential publication bias.

☒ Factors Related to the Magnitude of the Recidivism Effects

Multivariate analysis was conducted to identify the intervention characteristics most strongly associated with recidivism effects (see Lipsey, 2009, for a fuller account). Initial analysis focused on the relationship of methodological variables to the study findings. Those methodological variables showing significant relationships were carried forward in all analyses as statistical controls to minimize their influence on the substantive findings.

Characteristics of the Juveniles

It is widely presumed that specific services must be tailored to the unique treatment needs of very young offenders, girls, and different racial or ethnic groups. However, the meta-analysis showed that mean age, gender mix, and ethnic mix of the juvenile samples showed generally negligible relationships with the recidivism effects, indicating that intervention effects are relatively robust across these differences. The one characteristic of the juveniles receiving interventions that did show an overall relationship with recidivism outcomes was risk for delinquency. Interventions applied to high-risk delinquents, on average, produced larger recidivism reductions than for low-risk delinquents. However, the intervention effects were somewhat reduced for juveniles whose prior history included aggressive or violent offenses, albeit still significantly positive for reduced recidivism.

Juvenile Justice Supervision

Another question is whether certain programs are more effective in institutional settings while others are more effective in community settings. However,

with risk level statistically controlled, no differences were found in the recidivism effects for juveniles treated in the different juvenile justice supervision contexts—in the community or in secure facilities, or under probation supervision, or diverted. In other words, a program that is effective for juveniles treated in the community is also generally effective when it is administered to juveniles of similar risk in secure facilities and vice versa. This does not necessarily mean, however, that the overall outcomes of these respective placements are equal—only the average effects of the services themselves. Confinement often has other negative consequences, particularly for youth with mental health problems and when physical or sexual abuse and other forms of violent victimization are perpetrated by staff and other inmates.

Type of Intervention

It is no surprise that different types of interventions vary in their effectiveness for reducing recidivism. For example, cognitive-behavior therapy is widely acclaimed as an effective intervention. Indeed it is, but the good news is that other types of interventions were found to be effective as well. One important distinction has to do with the overarching philosophy of the program. “Philosophy” in this context means the global approach to altering juvenile behavior. Two broad program philosophies could be distinguished. The first features external control techniques for suppressing delinquency and includes three categories:

- Programs oriented toward instilling discipline (e.g., paramilitary regimens in boot camps).
- Programs aiming at deterrence through fear of the consequences of bad behavior (e.g., prison visitation programs such as Scared Straight).
- Programs emphasizing surveillance to detect bad behavior in the absence of treatment (e.g., intensive probation).

A contrasting philosophy involves attempts to bring about behavior change by facilitating personal development through improved skills, relationships, insight, and the like. This “therapeutic” philosophy includes the following types of programs:

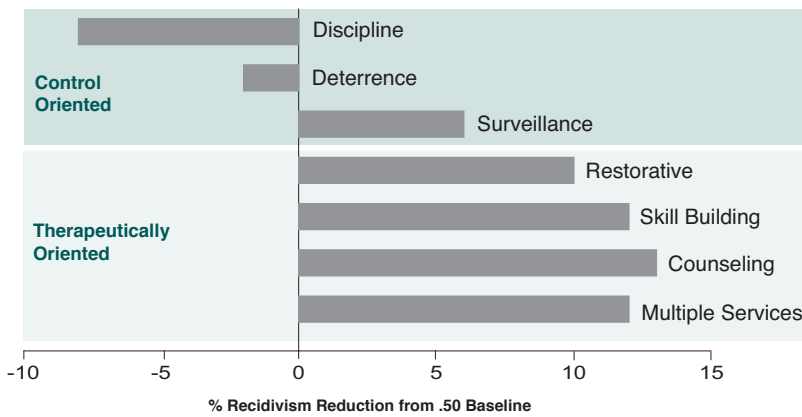
- Restorative (e.g., restitution, victim-offender mediation).
- Skill-building (e.g., cognitive behavioral techniques, social skills, academic and vocational skill building).
- Counseling (e.g., individual, group, family; mentoring).
- Multimodal or multiservice interventions designed to provide a package of multiple services tailored to the needs of individual juvenile offenders or groups of offenders.

When the effects on reoffense rates were compared for the programs associated with these two broad approaches, the programs with a therapeutic philosophy were notably more effective than those with a control philosophy. Figure 1 shows the effects for the program categories within each of these philosophies. The zero (0) point indicates no program effect while positive values represent reductions in

recidivism and negative values represent increases in recidivism. As can be seen, the programs in two of the control categories (discipline and deterrence) on average had negative effects.¹ The third category, programs relying mainly on surveillance, showed positive effects, but smaller ones than for any of the therapeutic program categories. This category includes mainly intensive probation programs, which often have significant counseling components by the probation officers. They may thus represent a mix of control and therapeutic strategies.

Figure 1

Mean Effects on Recidivism for the Major Intervention Approaches



For the purpose of guiding juvenile justice systems toward effective programs, the advice that follows from this portion of the meta-analysis is unequivocal. To optimize the effects on recidivism, programs in the therapeutic categories should be selected, not those in the control categories. It is especially important to avoid using discipline and deterrence programs that the research has shown typically have the unintended effect of increasing recidivism.

All the therapeutically oriented intervention approaches show average positive effects on recidivism that are nontrivial and generally similar across the different categories. Within each of these broad therapeutic intervention approaches, different generic intervention types that make up the category can be broken out. For example, Figures 2 and 3 show the average recidivism effects for the intervention types in the two largest program categories—skill building and counseling.

¹ All estimates of the reoffense effects have been adjusted for methodological differences between the studies, including how recidivism was measured.

Figure 2 shows that all the program types in the skill-building category had positive average effects, but behavioral programs (e.g., behavior contracting) and cognitive-behavioral programs had the largest effects. Figure 3 shows the average effects for the major generic types of counseling. Although they all show positive effects, the largest effects appeared for group counseling and mentoring programs. It should be noted that specific brand name model programs are embedded within many of these generic program types. These generally show positive effects on recidivism, as we would expect, but they do not necessarily show notably better

Figure 2

Mean Effects on Recidivism for the Generic Intervention Types within the Category of Skill Building Approaches

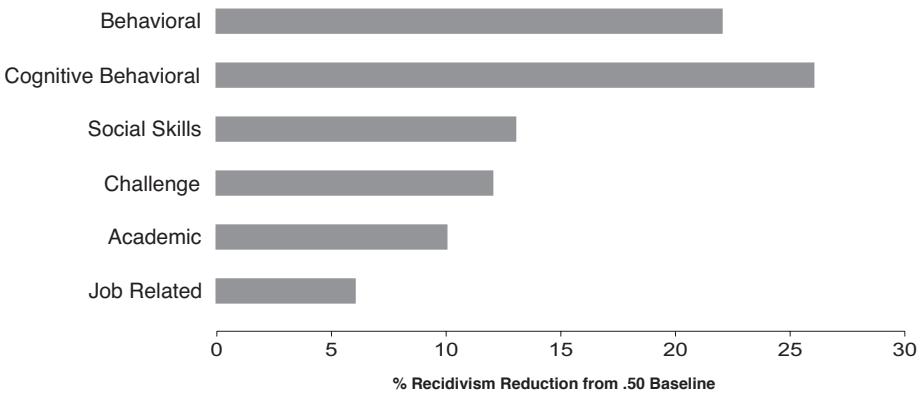
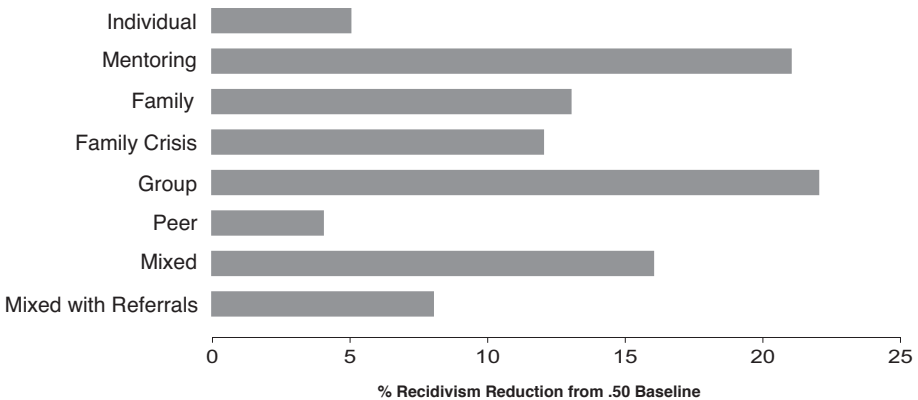


Figure 3

Mean Effects on Recidivism for the Generic Intervention Types within the Category of Counseling Approaches



effects than no-name programs of the same type. For example, Functional Family Therapy (FFT) and Multisystemic Therapy (MST) are both included in the generic program type labeled “family counseling.” The effects for those model programs, however, fall within the same range as the other family programs in this collection. Indeed, some no-name programs produced effects even larger than those found for the model programs.

Service amount and quality. Within a given intervention type, the recidivism effects varied in the expected manner with the amount of service provided—they were generally better with longer service duration and more contact hours up to a point of diminishing returns. The recidivism effects were also independently related to the quality of the implementation of the intervention. Higher quality implementation, as indexed by involvement of the researcher in the monitoring and quality control of the service delivery and no mention of implementation problems in the study report, was associated with larger recidivism reductions.

Other features of the intervention. The meta-analysis found no other general features of the interventions that were associated with the magnitude of their effects on recidivism. To be sure, there are other important features that were not sufficiently reported in studies to allow them to be coded and included in the analysis. For instance, we would expect the credentials of the therapists and service providers and, especially, the clinical quality of their services to influence the outcomes. Also, such characteristics of the juveniles as amenability to treatment and family support of the intervention services could well be influential. Meta-analysis is restricted to the information provided in the respective research reports and not all the information of potential interest and relevance was available. Nonetheless, each of the factors identified above showed relatively strong, statistically significant, independent relationships to recidivism effects. Though rather general, collectively they go a long way toward accounting for the variation in recidivism effects. As such, they provide a basis for generating guidelines about the characteristics that available research indicates interventions should have in order to produce strong positive effects on the recidivism of juvenile offenders.

☒ Best Practice Guidelines for Generic Intervention Types

The analysis results summarized above can be reformulated as guidelines for program providers about the profile of program characteristics expected to produce the largest effects on the recidivism of juvenile offenders. That advice can be stated in general terms as follows:

- Use therapeutically oriented approaches, not control-oriented ones.
- For the selected therapeutic approach, use one of the more effective intervention types within that category.
- For the selected intervention type:
 - Target high-risk juveniles; low-risk juveniles have little potential for recidivism.

- Provide an amount of service that at least matches the average in the supporting research for that intervention type.
- Implement the intervention with high quality; establish a treatment protocol and monitor service delivery for adherence to that protocol.

We have tested the utility of these simple meta-analytically generated practice guidelines by operationalizing them into the Standardized Program Evaluation Protocol (SPEP) program rating scheme that can be used by service providers and juvenile justice systems to assess their programs for juvenile offenders. The SPEP applies to any therapeutic intervention type for which there is a sufficient body of supporting research in our large meta-analytic database. The SPEP ratings are derived from data about the services the rated program actually provides, typically generated by a management information system maintained by the program provider or the juvenile justice system that uses the program.

Figure 4 illustrates the general form of the most recent SPEP rating scheme. The maximum number of points available in each category is proportionate to the strength of that factor for predicting recidivism effects in the meta-analysis. The services of the programs to be rated are classified into different types using descriptive information from the provider and a glossary of service descriptions derived from the associated research studies. Additional points are awarded if the program involves a therapeutic supplemental service that is also supported by favorable research in addition to the primary service. For example, when added as a supplemental service to a primary service of family counseling, a mentoring component for the juvenile may boost the effectiveness of the family counseling. The target values for treatment duration and hours of contact are set at the respective medians from the research on the intervention type being rated and vary for different intervention types. For manualized programs supported by research specifically on those programs, the amount of service targets specified by the developer are used instead. Data for rating the amount of service provided must come from a management information system or similar client tracking procedure that provides information for a sufficient number of juveniles served by the program in a recent period.

Service quality is the most difficult item to define empirically because research studies do not typically provide detailed information on this aspect. The quality indicators that were available for the meta-analysis tended to focus on monitoring implementation to ensure that it was done as intended. We, therefore, operationalized the rating of quality of service implementation as the presence of four elements:

- A written protocol or manual describing the intended services and how they are to be delivered;
- Staff delivering the service who have been trained in the program and associated protocol;
- Organizational procedures for monitoring adherence to the protocol and the quality of the service delivery; and
- Organizational procedures for taking corrective action when departures from the protocol or lapses in quality are identified.

Figure 4

Standardized Program Evaluation Protocol (SPEP) for Services to Juvenile Offenders Recalibrated version, 2012

	Points Possible	Points Received
Primary and Supplemental Service Types (Identified according to definitions derived from the research)		
Primary Service Type for Program Being Rated Group 1 services (5 points) Group 4 services (25 points) Group 2 services (10 points) Group 5 services (30 points) Group 3 services (15 points)	30	
Supplemental Service Type Qualifying supplemental service used: Yes (5 points) No (0 points)	5	
Quality of Service Delivery (Determined from a systematic assessment of the relevant features of the provider and provider organization)		
Rated quality of services delivered: Low (5 points) Medium (10 points) High (20 points)	20	
Amount of Service (Determined from data for the qualifying group of service recipients)		
Duration (Target number of weeks specified for each service type) % of youth who received at least the target weeks of service: 0% (0 points) 60% (6 points) 20% (2 points) 80% (8 points) 40% (4 points) 99% (10 points)	10	
Contact Hours (Target number of hours specified for each service type) % of youth who received at least the target hours of service: 0% (0 points) 60% (6 points) 20% (2 points) 80% (8 points) 40% (4 points) 99% (10 points)	10	
Risk Level of Youth Served (Determined from risk ratings on a valid instrument for the qualifying group of service recipients)		
% of youth with at least the target risk score set for the JJ system: 0% (0 points) 60% (15 points) 20% (5 points) 80% (20 points) 40% (10 points) 99% (25 points)	25	
Provider's Total SPEP Score	100	Insert Score

Finally, the risk level of the juveniles treated by the program must be determined from a valid risk assessment instrument administered routinely to each juvenile prior to treatment. As with the amount of service, the data needed to make the SPEP rating on this factor come from records for the juveniles served by the program during a recent period, with a minimum number of cases required before a rating can be made.

☒ Pilot Testing and Validation of the SPEP

The purpose of the SPEP is to translate information about the characteristics of effective programs found in the meta-analysis into a “best practice” guide that can be used by providers and juvenile justice agencies. As an evidence-based tool, the SPEP can be used to evaluate operational programs, designate them as evidence-based if they obtain high ratings, and guide improvement if they do not obtain high ratings. Demonstration projects with the SPEP have been conducted in the state juvenile justice agencies of North Carolina and Arizona, and another is underway in Tennessee. In all three implementations, providers and juvenile justice personnel have been generally accepting of the SPEP concept and its utility. The greatest practical difficulty has been obtaining appropriate data to support the SPEP ratings on a regular basis. All these states have management information systems that provide some of the data needed for the SPEP, but each has required some modifications tailored to the requirements of the SPEP. In addition, training of the information technology personnel was necessary for them to be able to extract the data required by the SPEP from their information systems and routinely produce the SPEP ratings for the programs of interest.

The only way to rigorously assess the validity of the SPEP for identifying effective programs would be to randomly assign juveniles to programs with high and low SPEP ratings and check whether recidivism rates were lower for the highly rated programs. That procedure was not feasible in any of the demonstration sites. Nevertheless, North Carolina data showed that the SPEP scores for rated programs were related to their recidivism rates (Lipsey, Howell, & Tidd, 2007).

In Arizona, however, we were able to conduct a more probing SPEP validation study (Lipsey, 2008). The Arizona Juvenile Justice Services Division has an especially good information system for recording the offense and juvenile justice history of the youth it serves. Records were pulled more than a year later for all the juvenile probationers who had been admitted during 2005–06 to one of the 66 SPEP-rated programs in the five pilot counties in which the SPEP was first implemented. There were 1,490 juveniles in this sample, all of whom had rearrest recidivism data for the six-month period after program completion and most of whom had 12-month recidivism data.

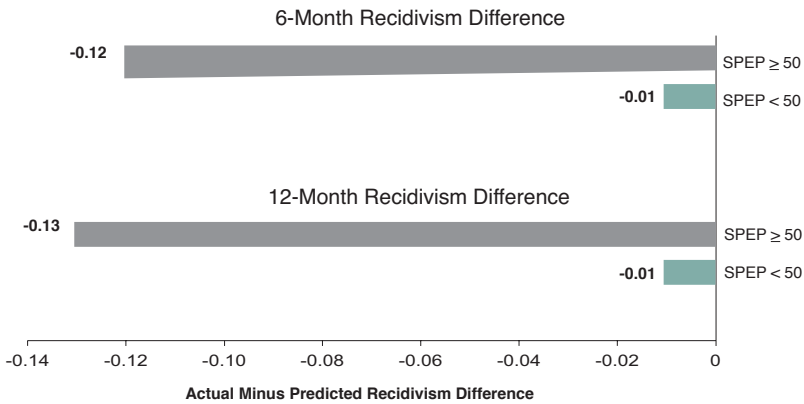
The juvenile justice history, risk assessment data, and juvenile demographic characteristics were used to predict the expected recidivism of the juveniles served by each of the 66 SPEP-rated programs. That estimate was then subtracted from

the actual recidivism rate for those juveniles. The difference thus took account of the different risk levels of the juveniles served by different programs so that they could be more fairly compared. If the SPEP score provides a valid index of the effectiveness of a program for reducing recidivism, the actual recidivism for juveniles served by highly rated programs should be lower than their predicted recidivism. Conversely, the actual recidivism for juveniles served by low-rated programs should be the same or more than their predicted recidivism.

Most of the 66 Arizona programs scored rather low on the SPEP—nearly 75% had ratings of 50 or less.² We should note that Arizona has a very good juvenile justice system and this result is not an indication of unusually poor quality programs but, rather, is typical of initial SPEP ratings. Further analysis of the SPEP scores revealed that most of the shortfall was due to service amounts below the target values. We then compared the actual and predicted recidivism rates for the juveniles served by the programs with ratings of 50 or less and those for programs scoring 50 or more. Figure 5 shows the results.

Figure 5

Difference between the Actual and Predicted Recidivism Rates for Juveniles Served by 66 Programs Scoring Above and Below 50 on the SPEP



For the six-month recidivism data, the recidivism rate for the juveniles served by the lower scoring programs was virtually the same as the rate predicted by their pretreatment history and characteristics—the difference was only .01, that is, a 1 percentage point difference in the recidivism rates. For the juveniles served by the higher rated programs, however, their recidivism rate was .12 less than what was predicted for them, that is, 12 percentage points lower than expected on the basis of their history and characteristics. The results were virtually the same for juveniles with 12-month recidivism data. A replication of this analysis with more

² Quality of service delivery in programs was not scored.

juveniles and SPEP-rated programs was conducted by the Arizona research staff after the SPEP had been rolled out statewide and showed similar results (Redpath & Brandner, 2010).

These recidivism studies provide promising indications of the validity of the SPEP for identifying effective programs and guiding improvement for ineffective ones. Our experience to date with the SPEP indicates that it can be used by service providers and juvenile justice agencies with only modest adaptations to their data collection and management information systems. Moreover, the logic of its ratings as direct representations of the predictive factors from the meta-analysis and the results of the North Carolina and Arizona recidivism studies give it credibility as a basis for judging the extent to which a program is supported by research evidence of effectiveness.

☒ Conclusion

There is no uniquely superior approach to evidence-based programming. Evidence for the effectiveness of individual local programs can come from direct evaluations of their specific effects. Evidence for the effectiveness of brand name protocol programs can come from research at one or more sites showing that programs that followed the protocol with fidelity had positive effects. And, evidence for the effectiveness of programs of a particular generic intervention type can come from meta-analysis of research showing positive effects for interventions of that type with certain specified best practice characteristics. The volume of meta-analyses has expanded enormously over the past few decades. Lipsey and Cullen (2007) recently conducted a comprehensive review of dozens of such systematic syntheses of the evaluations of programs for juvenile and adult offenders. Too few of these attempt to extract guidelines for practice, however, and more work along those lines would add greatly to their utility.

All of the approaches to evidence-based programs are useful for moving the juvenile justice field toward greater use of interventions supported by evidence that they are actually effective. Currently, however, evidence-based programs are almost universally taken to mean brand name protocol programs that have been certified by some authoritative group as having credible supporting evidence. When feasible, using such programs is advisable—if implemented with fidelity to the protocol, they have better odds of being effective than untested programs. However, there are many challenging issues associated with translating an evidence-based program into routine practice in a way that closely replicates the relevant circumstances of the original research. As a result, the desirable program effects on delinquency and subsequent offending found in the research studies may be attenuated when those programs are scaled up for general application.

Drawing on meta-analysis to identify the characteristics of effective programs of the type already in use in juvenile justice systems provides a complementary approach to the use of evidence-based protocol programs. This is an approach that

policymakers and practitioners can use to advance evidence-based practice when direct evaluation of local programs or implementation of brand name model programs is not feasible. Suppose, for instance, that a juvenile justice agency is making use of a local family therapy program that is believed to be well managed and generally effective. But no scientifically credible evaluation has been conducted to confirm its benefits for the juvenile offenders it treats, and resources for conducting such an evaluation are not available. To provide assurance that the agency's programs were evidence-based, this local program could be asked to adopt a family-oriented model program, such as Functional Family Therapy (FFT) or Multisystemic Therapy (MST) to replace their generic family therapy program.

Though that approach would indeed accomplish the objective of establishing evidence-based programming, it would also be disruptive to the local agency by requiring that its therapists be retrained to deliver the selected model program and adding the costs associated with that training and the acquisition and implementation of the program. This approach assumes that there is no evidence supporting the effectiveness of what the local program is already doing, hence the press to use a different program model. However, there are at least 29 controlled studies of the effects of family therapy on the recidivism of juvenile offenders, most of which do not involve brand name model programs but nonetheless perform as well. Moreover, the average effects shown in those studies are positive and large enough to have both practical and statistical significance. If the local family therapy program is conducted in a manner that matches the key features of the more effective programs represented in this research, then it can rightly be called evidence-based. If it does not quite match those key features, then its practices may be altered to provide a better match with much less disruption than adopting a new program model would require.

The SPEP was developed as a tool to crystallize the advice implied by the findings of the full body of research on effective programs for juvenile offenders into guidelines that could be easily understood and applied by program providers and juvenile justice agents. The SPEP makes it possible to assess the extent to which many programs already in place are consistent with research evidence about effectiveness for programs of that same type. Most important, perhaps, this tool provides guidance for improving programs that do not measure up well. Programs may be providing a type of service for which there is supporting research, but not doing so in a way that coincides with what the research indicates are the practices of the most effective programs of that type. There typically is much room for improvement. Identifying the areas with the greatest discrepancies between the program's practices and the research-based best practice guidelines gives programs a blueprint for improvement. If service providers and juvenile justice managers can collaborate in making collective improvements, they should be able to achieve recidivism reductions across multiple program types.

The prospect of making statewide program improvements appears to be within reach. The Comprehensive Strategy for Serious, Violent, and Chronic Juvenile

Offenders (Howell, 2003a, 2003b, 2009; Wilson & Howell, 1993) is a forward-looking administrative framework organized around risk management that promotes a statewide continuum of graduated sanctions and services that parallel offender careers. It incorporates best practice tools consisting of validated risk assessment instruments, reliable treatment needs assessments, a disposition matrix that guides placements in a manner that protects the public, protocols for developing comprehensive treatment plans that match effective services with offender treatment needs, and program quality assurance measures (Lipsey, Howell, Kelly, Chapman, & Carver, 2010).

The SPEP scheme, with its basis in the large body of research on delinquency interventions, also identifies programs for which the associated research shows negligible or adverse outcomes, such as those using external control techniques. Eliminating those from the repertoire can also help a juvenile justice system achieve better recidivism outcomes. Similarly, comparison with the body of research encompassed in the meta-analysis can identify programs in use for which there is no research. Such programs may be effective, or they may not, but there is no way to tell without appropriate research. The SPEP scheme facilitates the identification of such programs and alerts a juvenile justice system to their uncertainty. A local evaluation may then be conducted to provide the missing research, or they may be replaced with programs that already have research support. Best practice guidelines like the SPEP that are based on sound evidence can thus contribute in many different ways to improved juvenile justice programs. As Lipsey (2009) observed, "It does not take a magic bullet program to impact recidivism, only one that is well made and well aimed" (p. 145).

☒ References

- Howell, J. C. (2003a). Diffusing research into practice using the Comprehensive Strategy for Serious, Violent, and Chronic Juvenile Offenders. *Youth Violence and Juvenile Justice: An Interdisciplinary Journal*, 1(3), 219–245.
- Howell, J. C. (2003b). *Preventing and reducing juvenile delinquency: A comprehensive framework*. Thousand Oaks, CA: Sage.
- Howell, J. C. (2009). *Preventing and reducing juvenile delinquency: A comprehensive framework* (2nd Ed.). Thousand Oaks, CA: Sage.
- Lipsey, M. W. (1992). Juvenile delinquency treatment: A meta-analytic inquiry into the variability of effects. In T. D. Cook, H. Cooper, D. S. Cordray, H. Hartmann, L. V. Hedges, R. J. Light, T. A. Louis, & F. Mosteller (Eds.), *Meta-analysis for explanation: A casebook* (pp. 83–127). New York: Russell Sage Foundation.
- Lipsey, M. W. (1999a). Can intervention rehabilitate serious delinquents? *Annals of the American Academy of Political and Social Science*, 564, 142–166.
- Lipsey, M. W. (1999b). Can rehabilitative programs reduce the recidivism of juvenile offenders? An inquiry into the effectiveness of practical programs. *Virginia Journal of Social Policy and the Law*, 6, 611–641.
- 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, Center for Evaluation Research and Methodology. Available from: http://peabody.vanderbilt.edu/peabody_research_institute/pri_publications.xml.
- Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims and Offenders*, 4, 124–147.
- Lipsey, M. W., & Cullen, F. T. (2007). The effectiveness of correctional rehabilitation: A review of systematic reviews. *Annual Review of Law and Social Science*, 3, 297–320.
- Lipsey, M. W., Howell, J. C., Kelly, M. R., Chapman, G., & Carver, D. (2011). *Improving the effectiveness of juvenile justice programs: A new perspective on evidence-based practice*. Washington, DC: Georgetown University, Center for Juvenile Justice Reform.

- Lipsey, M. W., Howell, J. C., & Tidd, S. T. (2007). *A practical approach to evaluating and improving juvenile justice programs utilizing the Standardized Program Evaluation Protocol (SPEP): Final evaluation report*. Nashville, TN: Vanderbilt University, Center for Evaluation Research and Methodology
- Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders: A synthesis of research. In R. Loeber & D. P. Farrington (Eds.). *Serious and violent juvenile offenders: Risk factors and successful interventions* (Chap. 13, pp. 313–345). Thousand Oaks, CA: Sage.
- Redpath, D. P., & Brandner, J. K. (2010). *The Arizona Standardized Program Evaluation Protocol (SPEP) for assessing the effectiveness of programs for juvenile probationers: SPEP 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. Available from: http://peabody.vanderbilt.edu/peabody_research_institute/pri_publications.xml.
- Wilson, J. J., & Howell, J. C. (1993). *A comprehensive strategy for serious, violent and chronic juvenile offenders*. Washington, DC: U.S. Department of Justice, Office of Juvenile Justice and Delinquency Prevention.

James C. Howell is Managing Partner, The Comprehensive Strategy Group, Richmond, VA. Dr. Howell is former Director of Research and Program Development, Office of Juvenile Justice and Delinquency Prevention (OJJDP), U.S. Department of Justice. He has authored numerous publications on juvenile delinquency research, programs and strategies, including four books, the most recent of which is *Preventing and Reducing Juvenile Delinquency: A Comprehensive Framework*. He also is coauthor with John J. Wilson of the OJJDP Comprehensive Strategy for Serious, Violent, and Chronic Juvenile Offenders.

Mark W. Lipsey is the Director of the Peabody Research Institute and a Research Professor at Vanderbilt University. He specializes in program evaluation and field experimentation methodology, with a focus on programs for at-risk children and youth. His current research involves the study of risk factors and effective intervention for antisocial behavior and delinquency, early childhood education programs, and issues of methodological quality in program evaluation. Prof. Lipsey's research has been funded by major federal agencies and private foundations and recognized with awards from several professional organizations. He serves as editor or editorial board member for a number of criminology, evaluation research, and meta-analysis journals and has published books, journal articles, book chapters, and technical reports in those topic areas.