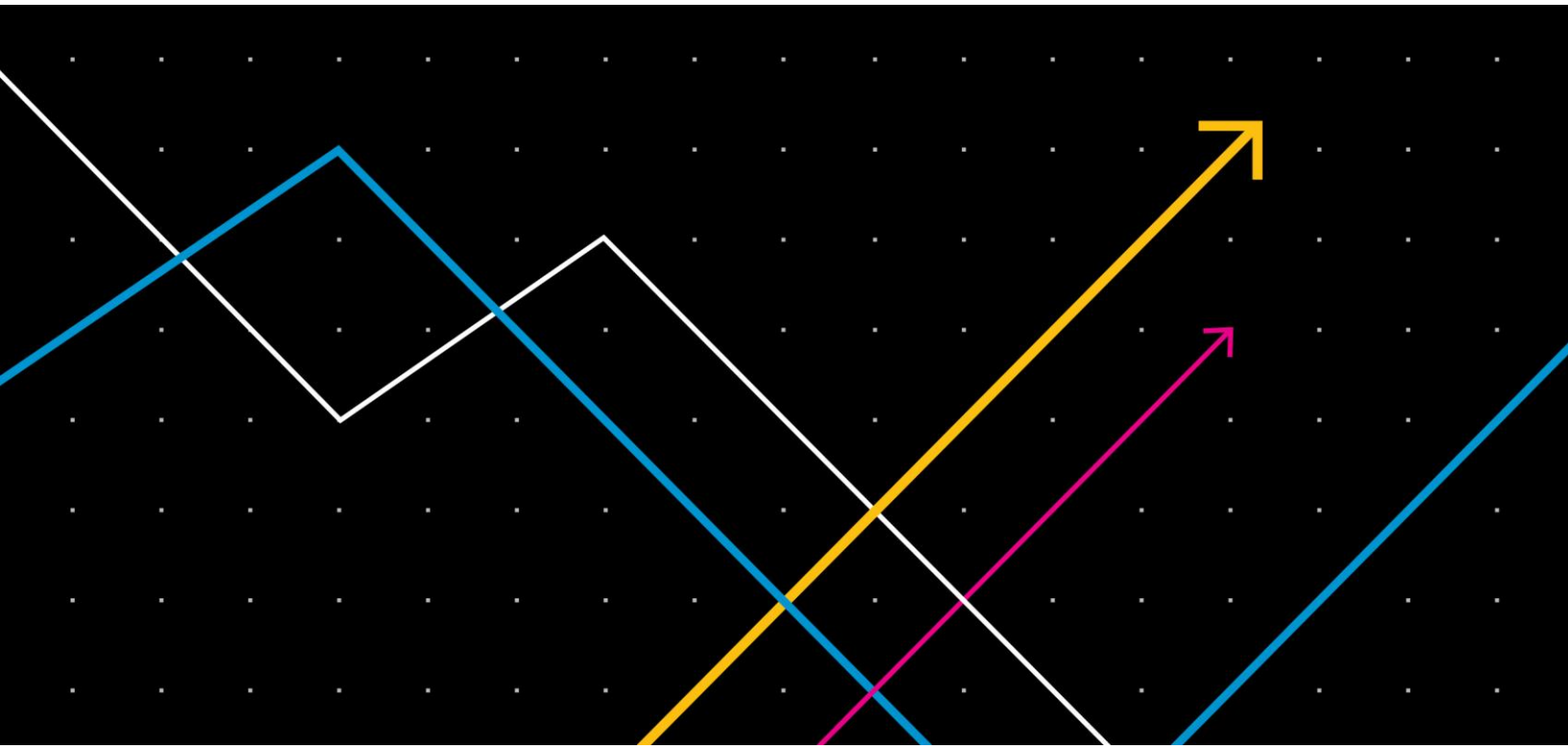


RESEARCH REPORT

Implementing the SPEP™

Lessons from Demonstration Sites in OJJDP's Juvenile Justice Reform and Reinvestment Initiative

Akiva Liberman *Jeanette Hussemann*

December 2016

URBAN INSTITUTE · ELEVATE · THE · DEBATE

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

ABOUT THE URBAN INSTITUTE

The nonprofit Urban Institute is dedicated to elevating the debate on social and economic policy. For nearly five decades, Urban scholars have conducted research and offered evidence-based solutions that improve lives and strengthen communities across a rapidly urbanizing world. Their objective research helps expand opportunities for all, reduce hardship among the most vulnerable, and strengthen the effectiveness of the public sector.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

Contents

Acknowledgments

This report was funded by Grant No. 2012-PF-F2-K001 awarded by the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, US Department of Justice. Points of view expressed in this document are those of the authors and do not necessarily represent the official position or policies of OJJDP or the US Department of Justice. We are grateful to them and to all our funders, who make it possible for Urban to advance its mission.

The views expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders. Funders do not determine research findings or the insights and recommendations of Urban experts. Further information on the Urban Institute's funding principles is available at www.urban.org/support.

1. Introduction

At the end of 2012, the Office of Juvenile Justice and Delinquency Prevention (OJJDP) launched the Juvenile Justice Reform and Reinvestment Initiative (JJRRI) in three demonstration sites in Delaware, Iowa, and Milwaukee County, Wisconsin. The goal of JJRRI was to bring evidence and best practices to bear on juvenile justice operations. The primary vehicles used in JJRRI were the development of dispositional matrices to provide evidence-based recommendations concerning dispositional options and the implementation of the Standardized Program Evaluation Protocol (SPEP™) rating system to assess and guide improvements in the effectiveness of programs delivered to juvenile justice youth. Together, these tools were intended to increase the effectiveness and efficiency in the use of juvenile justice resources.

Concurrent with the implementation of JJRRI at the three demonstration sites, the Urban Institute conducted a process and outcome evaluation of the reform initiative. The goals of the evaluation included understanding whether the implementation of JJRRI improved the quality and effectiveness of juvenile justice programming. More specifically, the process evaluation aimed to understand the implementation of the SPEP™ at the demonstration sites, including implementation requirements, successes, and challenges, and implications for sustainability and replication.

This report focuses specifically on the implementation of the SPEP™ rating system at the three JJRRI demonstration sites. Two subsequent reports will address local validation of SPEP™ ratings and overall JJRRI implementation.

Findings are based on data collected between 2012 and 2015. Data collection included annual visits to each site with technical assistance (TA) providers, observation of on-site trainings, and in-depth annual telephone interviews with stakeholders to monitor progress and assess stakeholder perspectives. Interviewees included juvenile justice administrators, program providers, court workers, contractors, data managers, and other staff implementing the SPEP™. Additional information was collected through written reports and narratives provided by the JJRRI sites, as well as participation in regular calls with the sites, funders, and TA providers.

This report describes implementation of the SPEP™ as an important element of the JJRRI work at the three demonstration sites. Chapter two briefly discusses other key components of JJRRI and the context in which the SPEP™ was implemented, including a short description of the demonstration sites. Chapters three and four describe the SPEP™ and the drivers of SPEP™ implementation. Chapter five

discusses how implementation proceeded in the JJRRI demonstration sites. Finally, chapter six discusses lessons learned.

2. JJRRI

JJRRI is a demonstration program aimed to reduce recidivism by improving the services for youth in the justice system and thereby using evidence to increase both the effectiveness and efficiency of the use of juvenile justice resources. OJJDP funded three diverse JJRRI demonstration sites at the end of 2012:

- Delaware: Implementation of the SPEP™ in Delaware was coordinated by the Division of Youth Rehabilitative Services (DYRS), a division of the Department of Services for Children, Youth and their Families (DSCYF). DYRS is responsible for providing supervision to preadjudicated and adjudicated youth throughout Delaware.
- Iowa: JJRRI was implemented in the first, third, and sixth Iowa judicial districts with strong support from the three chief juvenile officers in those districts. Chief juvenile officers oversee juvenile program and service contracts and community supervision of youth formally charged with a delinquent act in their respective districts. Implementation was coordinated by the Iowa Division of Criminal and Juvenile Justice Planning (CJJP), a division of the Iowa Department of Human Rights that serves as the Iowa Statistical Analysis Center and is located in the executive branch.
- Milwaukee County, Wisconsin: JJRRI in Milwaukee County was coordinated by the Milwaukee County Delinquency and Court Services Division (DCSD). DCSD is responsible for the administration of non-judicial operational services, intake and probation services, operation of a 120-bed detention facility, and provision of purchased services. DCSD serves youth from referral though the end of court–imposed supervision.

Appendix A provides additional information about the sites. How each site organized and implemented their SPEP™ work is discussed in the following chapters.

It is important to understand how the SPEP™ fits into the broader goals of JJRRI and how SPEP™ work occurred alongside other reform efforts of the initiative. Although implementation of the SPEP™ was central to the work of JJRRI, the initiative also focused considerable attention on developing dispositional matrices as tools to guide juvenile justice decisionmaking and ensuring that risk assessment data, used in both the SPEP™ and dispositional matrices, were valid at the local level.

Dispositional Matrices

Dispositional matrices typically combine information about risk of recidivism, determined by a validated risk assessment tool, with current case information. Together, these considerations generate recommended dispositional options. For example, violent felony charges on the current case combined with medium risk of recidivism might lead to a recommendation for out-of-home placement, followed by intensive supervision, while misdemeanor charges and low risk might lead to a recommendation for court diversion. Dispositional matrices are an important tool for generating evidence-based recommendations concerning the effective use of dispositional decisions and appropriate placement. Over the course of JJRRI implementation, the demonstration sites worked with local juvenile justice stakeholders, including the judiciary, to develop local dispositional matrices to inform juvenile justice decisionmaking.

In JJRRI, a dispositional matrix that has been developed and tested in Florida was generally used as a model. Florida spent considerable time developing and validating its tool. Early research on Florida's dispositional matrix has shown that when dispositions are outside the range of recommended dispositions, recidivism is typically higher than when dispositions were within the recommended range (Boglivio, Greenwald, and Russell 2015). See appendix C for a copy of Florida's dispositional matrix.

Risk Assessment

Valid and reliable risk assessments are a prerequisite to implementing both the SPEP™ and dispositional matrices and to promoting appropriate service placement for juvenile justice youth. Thus, an important goal of JJRRI was to promote the reliable use of a risk assessment that had been locally validated. In practice, categorical levels of risk (e.g., low, medium, or high risk) are used in a dispositional matrix, with different dispositional options being recommended for youth with different risk levels for recidivism. Similarly, the SPEP™ rating system uses these categorical levels of risk, which are generally produced from continuous risk scores, and the cutpoints for differentiating levels of risk can vary locally. Thus, one important aspect of local validation of an established risk assessment system involves setting appropriate *local* cutpoints for levels of risk that take into account the distribution of recidivism risk among the local juvenile justice population. Because this work can be time and data intensive, JJRRI provided considerable TA to the demonstration sites to work through risk assessment data issues. The SPEP™ requirements for risk assessment are discussed below in chapters three and four, and the sites' experience with its implementation is discussed in chapter five.

Reducing Racial and Ethnic Disparities

Another important goal for JJRRI was to reduce racial and ethnic disparities (RED). Thus, one of the aspirations for the use of evidence-based tools at disposition (i.e., a dispositional matrix) and for the improvement of services (i.e., the SPEP™) was to reduce RED in these aspects of the system. Because of the centrality of valid risk assessment, this also meant that it was important to validate that the risk assessments in use at JJRRI sites were valid both within and across racial and ethnic groups (see Baird et al. 2013). In addition, sites began exploring whether there seemed to be differences in the services offered to comparable youth of differing race and ethnicity. TA provided to JJRRI sites was essential in trying to identify and address any such RED issues.

Although the work of JJRRI was broader than the SPEP™, the SPEP™ was its most prominent implementation component. As the SPEP™ is a distinct tool that is being used to increase the effectiveness of juvenile justice services, this report focuses specifically on the implementation requirements and challenges of the SPEP™.

3. The SPEP™

The SPEP™ provides local jurisdictions with an evidence-based tool to rate services delivered to youth in the juvenile justice system based on their potential to reduce recidivism. Developed by Mark Lipsey of Vanderbilt University's Peabody Research Institute, the SPEP™ is based on the quantitative synthesis of evidence from more than 600 controlled evaluations of the effectiveness of interventions in reducing recidivism among justice-involved youth (Lipsey 2009). Based on this meta-analysis, Lipsey identified four basic characteristics of effective juvenile justice services and used them to develop the SPEP™ tool: the type of service, the quality of the service, service dosage, and the risk level of youth who receive the service. Taking into consideration these four elements of effective services, the SPEP™ provides local jurisdictions with the ability to rate juvenile justice services in accordance with evidence on what services are shown to effectively reduce juvenile justice recidivism.[1]

The SPEP™ is generally used as part of a continuous quality improvement (CQI) cycle of ratings and quality improvements; the SPEP™ ratings guide program improvement, which should lead to improved ratings in the next round. The SPEP™ is intended as a vehicle for quality improvement at the level of both individual programs and the juvenile justice system. At the program level, the SPEP™ ratings serve as feedback to programs regarding areas of possible improvement. If most of the programs being used by a jurisdiction are rated on the SPEP™, then an examination of SPEP™ ratings across the entire system can be used to assess the current array of programs. This may show that the array of programs could be improved or expanded to provide more diversity in program options, that the wrong youth are being referred to programs, and that programs are not generally contracting for effective doses of services to be delivered. Such improvements at the system level are sometimes referred to as improving the "system alignment."

Below, the four elements of the SPEP™ are briefly reviewed. Two elements are assessed at the program level, and two are assessed at the individual client level and then aggregated to the program level.[2] Although the SPEP™ is intended to guide a cycle of program ratings, collecting necessary data and completing a first round of SPEP™ ratings requires considerable time, effort, and TA.

Program-Level Elements of the SPEP™

Two SPEP™ elements are assessed at the level of the service. The first SPEP™ element concerns the type of service being delivered. The research evidence indicates that some types of services are more

effective in reducing recidivism than others. Programs with a therapeutic orientation (e.g., counseling and skill-building programs) are much more effective at reducing juvenile recidivism than programs with a control orientation (e.g., boot camps and "scared-straight" programs). Some types of therapeutic programs have been found more effective than others. Rating this SPEP™ element involves categorizing each service that is being delivered into the types of services in the research literature. Note that when one program delivers multiple services to the same youth, such as individual counseling and social-skills training, each service must be rated separately on the SPEP™ because each service has different potential to reduce recidivism.

The research on which the SPEP™ is based has found that service effectiveness does not depend on whether services are delivered in a community setting or a residential setting (Lipsey 2009). Therefore, both community and residentially delivered services can be rated on the same criteria. Because the youth in a residential setting are often delivered "bundles" of services, more work tends to be involved in unbundling residential services before those services can be assessed and rated with the SPEP™.

The second SPEP™ element concerns the ability of the program to deliver its service consistently and reliably, which is referred to as the "quality of service delivery." This element focuses on whether programs are structured to promote consistent and reliable delivery of the intended services, such as whether programs have manuals, initial and continuous staff training, program monitoring, and protocols for remedial action to correct program gaps or drift. This element of the SPEP™ is the least standardized, and the SPEP™ developers have worked with sites to customize how this element can be rated given local circumstances (Lipsey et al. 2010).

These two elements of the SPEP™ may require considerable TA but are not especially data intensive.

Client-Level Elements of the SPEP™

The other two SPEP™ elements are more data intensive because they require data at the level of the individual clients. These elements include the dosage of service provided to each youth and the risk level of youth who receive the services.

Service dosage data includes detailed information on the amount of a service received by each youth. The research that underlies the SPEP™ is based on evaluations of recidivism reduction for services delivered with some dosage (e.g., 20 hours of service); when the same service is delivered at a

lower dose, the prior evidence of effectiveness does not generalize. Dosage targets for the SPEP™ are based on both the number of contact hours for each youth (e.g., 2.5 hours per week) and the duration in which services are delivered (e.g., 12 weeks). The SPEP™ rates a program for the percentage of its juvenile justice clientele who meet dosage targets for the service.

When youth receive multiple services from one program, the dosage data must be established separately for each service that is rated on the SPEP™; the contact hours are not combined. For example, if Alex received 15 hours each of social skills training (dosage target = 24 contact hours) and family counseling (dosage target = 30 contact hours), neither service was delivered at target levels. Thus, despite receiving 30 total hours of services, Alex would not be counted as having received the effective necessary dosage for either service. This means that records of client contacts must be specific to each service that is being rated on the SPEP™.

The final element of the SPEP™ concerns which youth receive a service. Research consistently finds that programs more effectively reduce recidivism when delivered to youth at higher risk of recidivism (Lipsey 2009); this is often referred to as the "risk principle" (Andrews, Bonta, and Hoge 1990). Basically, recidivism cannot be reduced much for youth with little risk of recidivism. Therefore, services, especially intensive services, are not indicated for low-risk youth.[3] Therefore, in order for the SPEP™ assessment to be conducted, a system must assess a youth's risk of reoffending—using a validated risk tool—*before* services are delivered.

Finally, it is important to note that the SPEP™ does not rate how youth are matched to services based on their specific needs. Although it is generally accepted that not all youth need the same services and treatment matching is commonly believed to be important for program effectiveness, Lipsey's meta-analyses of prior program evaluations have not shown a pattern in which better matching is associated with more effective recidivism reduction for juvenile (Howell and Lipsey 2012). As a result, beyond targeting youth at high risk of recidivism, the SPEP™ does not rate matching youth to services.4 Relatedly, the SPEP™ does not currently have separate ratings for services for specialized populations, such as sex offenders or youth with mental illness.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

4. Key Drivers of SPEP™ Implementation

Although the SPEP™ tool is fairly simple in conception, implementing the first round of ratings in a jurisdiction is generally a complex enterprise. For example, JJRRI demonstration sites generally took two to three years to conduct one round of SPEP™ ratings. Why should this process take several years?

This chapter addresses this question by describing what is involved in implementing the SPEP™, including three general classes of interdependent "drivers" of SPEP™ implementation. This report refers to key systems of support that facilitate successful implementation of the SPEP™ as "drivers" of implementation (Metz and Bartley 2012). Drivers include both absolute requirements (e.g., types of data) and other supports without which implementation is generally not completed successfully.

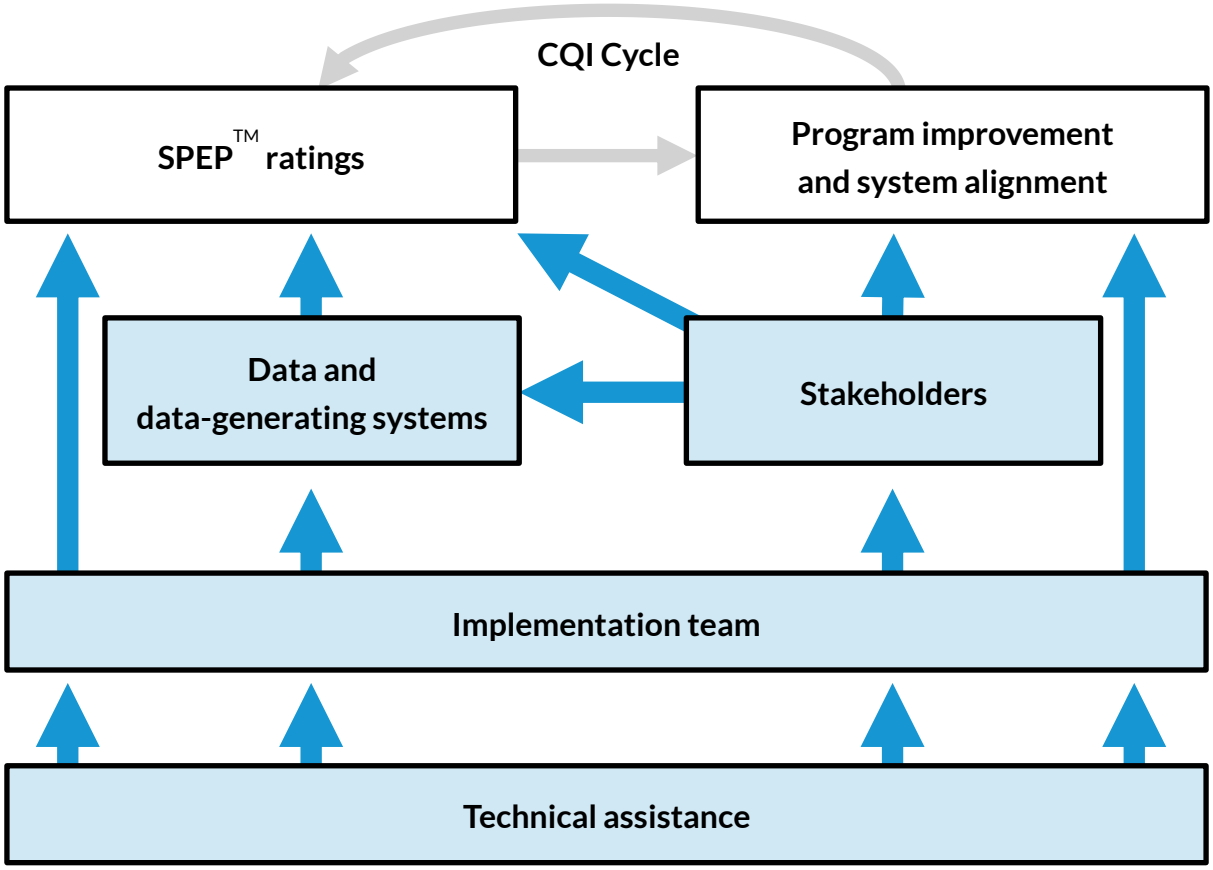
The evaluation identified four key drivers to SPEP™ implementation. First, the SPEP™ requires data systems containing the information needed for SPEP™ ratings, as well as systems to reliably generate that data. Although jurisdictions may believe that these systems are in place before embarking on the SPEP™ rating process, the attempt to conduct a first round of rating generally uncovers inadequacies in these systems that must be addressed. Improving data or data-generating systems requires considerable stakeholder support, which is the second implementation driver. Support is required from people and organizations with stakes in the data and data-generating systems, as well as those with stakes in the CQI effort more broadly, which includes those with stakes in the program and systems that the SPEP™ ratings are intended to help improve.

Third, implementation requires the coordinated efforts of a team of people. In JJRRI, this work was organized through an implementation team led by a JJRRI program manager. The team included a subset of juvenile justice stakeholders from both local and state levels who supported the goals of JJRRI and facilitated buy-in of SPEP™ implementation among juvenile justice administrators and service providers. Finally, implementing the SPEP™ required considerable TA to support technical aspects and build and buttress the stakeholder support for the entire CQI and juvenile justice reform effort. This TA was generally delivered to the implementation team as a first matter and to other stakeholders as appropriate.

These implementation drivers and how they support each other are illustrated in figure 1.

FIGURE 1

Implementation Drivers for the SPEP™



Data Systems and Data-Generating Systems

Each element rated on the SPEP™ requires data. For SPEP™ program-level elements, which require only one datum for each service, rating these elements may take time, effort, and TA, but they do not impose a large burden on data or data systems. In contrast, the other two data elements need to be assembled at the individual-client level and thus have substantial data requirements. For each service, the client-level elements are based on a cohort of youths identified by their entry into the service during a well-defined period of time (e.g., a 12-month period, starting on January 1, 2013). Of course, these SPEP™ measures more accurately describe a service when they are based on larger numbers of clients. The SPEP™ developers have set a minimum threshold of a cohort of 10 youths per service that is rated on the SPEP™. For small services, it may take considerable time to accrue enough clients for a reliable SPEP™ rating.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

The SPEP™ ratings are based on the percentage of clients who exceed thresholds of risk and required dosage. Thus, these two elements require not only well-developed management information systems or other data systems to store and update this data systematically, but also systems to generate those data. For risk data, a risk assessment system is needed, including a risk assessment tool, as well as policies and practices concerning who conducts risk assessment, when, and how often. For dosage data, the data in some cases may be produced by existing billing or other systems, but in other cases this may require the establishment of new systems and procedures for recording client contact.

Risk data must be relatively complete for the period during which a program is rated on the SPEP™. If risk assessment is completed only on a minority of cases, then the youth with risk data cannot be taken as representative of the clientele as a whole. The SPEP™ developers require individual risk data for 80 percent of youth in a service, based on risk assessment conducted within 90 (preferably 60) days preceding initiation of the service, in order to produce a "full SPEP™ rating" of that service.

The collection of individual-level data elements has proven most likely to require modification of existing systems. Risk data has often been found to be incomplete, not timely, or suboptimal in some other way. This, in turn, leads to a reexamination of the risk assessment system in use, who conducts the risk assessment and when, and the data system that captures the risk data. In JJRRI sites, risk data was incomplete in various ways, as discussed in the next chapter.

Whether client dosage data is systematically recorded often depends, in part, on the nature of the contract under which services are delivered, who pays for services, and if there are intermediary organizations involved, as well as the nature of the billing system.

Stakeholder Support

Each of the SPEP™ elements has stakeholders whose support is needed for the effort to succeed and whose resistance can impede implementation. Table 1 summarizes possible stakeholders for each SPEP™ element, for the SPEP™ rating itself, for program improvement, and for system alignment. Support from all of these stakeholders is important early in the process.

Stakeholders in the data include both data owners and those who generate data. For the two elements involving extensive client-level data, namely risk and dosage, the owners of data systems are especially important. Because the structures of juvenile justice systems vary at state and local levels, the particular stakeholders vary by jurisdiction. As one example, probation officers commonly conduct

risk assessments. Whether juvenile probation is an arm of the court or the executive affects whether the chief judge or the governor is the superordinate data owner of risk assessment data systems.

For data-generating systems, some stakeholders have stakes in the data besides the SPEP™ ratings. For example, improving the systematic assessment of risk before service initiation may require the support of probation staff, who often conduct risk assessment. Whether probation is part of the juvenile corrections agency, in turn, may change the stakes that the juvenile corrections agency has in the use of risk assessment. Moreover, changes to when or how risk assessment is conducted may require the support of judges, prosecutors, and the defense bar, who may have stakes in when and how risk data is collected and made available. The people and organizations whose activities are described by the data produced also have stakes in that data. Given that data is being collected to guide program improvement, stakeholders include the service providers and others who must support or implement improvements to particular programs. Finally, because the SPEP™ ratings are intended to guide system alignment efforts, including possible changes to referral patterns and contracts, stakeholders include actors and organizations that implement, fund, or authorize system-level changes.

In all JJRRI sites, considerable time and effort were involved in messaging around SPEP™ ratings to make feedback from the SPEP™ constructive. The effort was consistently described as a source of evidence-based feedback to help programs improve their effectiveness, rather than a "gotcha" exercise to identify and sanction poor programs. The SPEP™ developers stressed that jurisdictions need an array of different types of effective services and not all youth need the type of services with the highest potential SPEP™ scores. For example, in the SPEP™ rating scheme, cognitive behavioral therapy (CBT) is the service type with the highest potential SPEP™ score, but CBT is not appropriate for all juvenile justice youth. As part of communicating the implications of the SPEP™ ratings for possible program improvement, the SPEP™ developers have also developed Program Optimization Percentage (POP) ratings to indicate the program's room for possible improvement. Some services may be contracted for a particular type of service or to serve medium- or low-risk youth. Under those circumstances, the service cannot achieve a SPEP™ rating of 100. The POP ratings represent the SPEP™ as a percentage of the service's highest potential score.

More generally, sites and TA providers need to make considerable efforts to prevent misuse or misunderstanding of SPEP™ ratings. This also involves considerable deliberation about when and to whom SPEP™ ratings (or subratings) should be released.

In sum, a general implication is that using the SPEP™ in a CQI effort involves a large reform to the juvenile justice system, so that many key juvenile justice agencies and decisionmakers are stakeholders in the process.

TABLE 1

Stakeholders for the SPEP™ Continuous Quality Improvement Process

SPEP™ elements	Stakeholders		
	SPEP™ data and data generation Including data owners and producers; organizations and programs that receive SPEP™ ratings and reports	**Program improvement** Including parties responsible for program improvement	**System alignment** Including funders, authorizers, and implementers
Service type Therapeutic services	Program staff to produce manuals and service descriptions; contracting to require and store service descriptions	Program staff to revise or produce manuals and service descriptions; contracting to require and store service descriptions	Service providers, contracting, funders, and intermediaries; legislature
Quality of service delivery Quality assurance data and systems	Service providers to produce service and training manuals and monitoring and corrective actions	Service providers, contract monitors, and intermediaries to revise and/or monitor service quality	Service providers, contracting, funders, and intermediaries; legislature
Risk assessment Assessing risk	Probation offices and officers to complete assessments		
Risk data and systems	Probation, court, and information systems departments	Juvenile courts, including judges, prosecutors, defense, and probation, to appropriately place youth	Juvenile courts, including judges, prosecutors, defense, and probation, to appropriately place youth; legislature
Dosage Service participation data and systems	Program staff to provide and document service participation; funders of services for juvenile justice youth (including juvenile justice, mental health, and child welfare agencies); contracting and information systems departments	Program staff to modify dosage; funders to support increased dosage (including juvenile justice, mental health, and child welfare agencies); intermediary agencies; contracting	Funders and legislature

Implementation Team

In JJRRI, implementation teams were organized at each site to accomplish much of the SPEP™ work around identifying and assembling the necessary data and to bring an important set of stakeholders together. In view of the diverse set of stakeholders needed and the extended period of time involved in getting SPEP™ ratings implemented, this structure was seen as an important vehicle for keeping the work coordinated and moving forward and for keeping a set of stakeholders engaged in a sustained manner. Implementation teams involved a combination of staff who manage the data needed for the SPEP™; administrators and other decisionmakers; representatives from other stakeholder constituencies, such as service providers or juvenile probation officers; and others. The next chapter briefly describes the makeup of the implementation teams at each site.

In JJRRI, the implementation teams were headed by the program managers, whose positions were funded by the JJRRI grants. Because keeping the work organized involved considerable staff time, this proved an important support for the SPEP™ process. In addition, TA providers helped the implementation team set up action plans to help organize and monitor the work.

Technical Assistance

Effective implementation of the SPEP™ rating and CQI systems requires considerable TA. This TA was infused throughout JJRRI's implementation of the SPEP™, as is described in more detail in the next chapter. A brief sketch of the types of TA that were involved is provided here.

To conduct the first round of SPEP™ rating required several types of TA. Two were technical in a narrow sense: TA around the SPEP™ rating itself and TA around the data and data-generating systems that are needed for the SPEP™. In JJRRI, as a first matter, the implementation teams were the primary recipients of the TA necessary to directly implement the SPEP™.

In addition, considerable TA was needed to generate and buttress the support of the many stakeholders for the SPEP™-based CQI process. To provide the latter required presentations to multiple audiences concerning the evidence base for the SPEP™ and, more broadly, for evidence-based juvenile justice reform. The implementation teams were a primary audience for this TA and conduits to other key stakeholders whose support and buy-in were needed for the SPEP™ process.

Finally, after the first round of SPEP™ ratings, JJRRI also involved the provision of considerable TA to help guide program improvement efforts based on initial SPEP™ ratings. This was an important

aspect of the SPEP™ implementation. Although some CQI improvement efforts expect the feedback recipients to use the feedback themselves to devise their own program improvement efforts, prior experience with the SPEP™ found that sites sought additional TA for program improvement planning. This TA helped sites capitalize effectively on the feedback that the SPEP™ provided and build and institutionalize a CQI process that would be sustained beyond the grant.

5. SPEP™ Implementation in JJRRI Sites

At the end of 2012, OJJDP funded three demonstration sites to implement JJRRI and engage in SPEP™ work. The demonstration sites differed in important ways, as described below, and thus provided a suitable platform for learning about implementation of the SPEP™. After providing a brief overview of each demonstration site, the implementation experiences in the JJRRI sites during their first three years are discussed, with particular attention to the challenges and barriers encountered by the sites in relation to the SPEP™ implementation drivers—data systems, stakeholder support, implementation team, and technical assistance.

SPEP™ at the JJRRI Demonstration Sites

Delaware

Delaware focused its SPEP™ work on community-based services that were being provided to higher-risk youth and decided not to rate services that were provided only to low-risk youth. In its grant application for JJRRI, Delaware proposed focusing its SPEP™ ratings on newly contracted programs that the state had not yet evaluated or audited.

By policy, risk assessment was administered to youth following adjudication and was limited to youth who received community supervision. Because risk data are central for the SPEP™ ratings, this precluded SPEP™ ratings for services provided to youth in residential placements.[5]

At the time of SPEP™ implementation, Department of Youth and Rehabilitative Services (DYRS) providers used the Family and Child Tracking System (FACTS) to track juvenile cases and services, which allowed DYRS staff to access juvenile justice information and youth assessment and service information in other sister divisions (the Division of Family Services, Division of Prevention and Behavioral Health Services, and the Division of Management Support Services). It did not include risk data, which was stored in a separate database.

Delaware's implementation team included state and local juvenile justice representatives. In addition to the program manager, key members included the newly appointed DYRS director, the chief of community services, and the regional manager of pretrial and low-level juvenile services, as well as representatives from the juvenile court. Several IT and data specialists participated in the team, including the director of the Criminal Justice Council Statistical Analysis Center, to support improved data infrastructure for the collection of SPEP™ data elements. Several administrators of community-based service providers that were to be included in the first round of SPEP™ ratings also participated and "represented" youth services provided in all three counties in Delaware.

Iowa

Iowa implemented the SPEP™ with both community-based and residential programs. Community-based programs were contracted by each judicial district, which received funding from Iowa Division of Criminal and Juvenile Justice Planning (CJJP) and the Department of Human Services (DHS). Contracting mechanisms varied by district and program. Residential programs were funded through CJJP, DHS, and judicial districts.

Risk information was stored in the Iowa Court Information System (ICIS), which is a statewide data system that stores juvenile and criminal justice case information, including court services, case processing, financial reporting, and appellate record review. The system was regularly updated by juvenile court staff in all judicial jurisdictions.

SPEP™ implementation involved statewide residential programs and local community-based services in three judicial districts. Iowa's implementation team was headed by CJJP staff, including the program manager, division director, and IT specialist. They also led SPEP™ data collection at state residential programs. The chief juvenile court officers (CJCOs) of the first, third, and sixth judicial districts participated in the implementation team and facilitated SPEP™ data collection in their local districts. CJCOs also selected local juvenile justice workers, including judges, contractors, and community service providers to participate in the implementation team.

In Iowa, much of the work of the juvenile justice system is led by the CJCOs, who are in the judicial branch. With the support of CJCOs from the three participating judicial districts, CJJP, the primary grant recipient for JJRRI located in the executive branch, served as the SPEP™ coordinator. Because CJJP also serves as the state's criminal justice Statistical Analysis Center, it also served as the data collection and reporting agent. This arrangement meant that coordination and collaboration was key to

implementation in Iowa. It also posed a challenge to sustainability past the grant period, in that funding to sustain staff time at CJJP was dependent on a combination of the executive and legislative branches, to support work in the judicial branch.

Milwaukee County, Wisconsin

In Milwaukee, programs involved in the SPEP™ ratings included programs funded directly through Milwaukee County Delinquency and Court Services Division (DCSD) or through Wraparound Milwaukee, a managed care program operated by the Behavioral Health Division to serve clients with a DSM-IV diagnosis. Only a minority of Wraparound Milwaukee clients are juvenile justice youth; most services are funded through DCSD, child welfare, mental health, and Medicaid capitation. Residential and community-based programs from each agency were included in the SPEP™. Coordinated and individual juvenile justice programming was typically funded through fee-for-service or unit-price contracts.

DCSD relied on multiple data systems to track juvenile justice information. Youth case processing information was stored in the Juvenile Information Management System, which tracked youth from court referral through the end of court–imposed supervision. Case management and service information for youth receiving services through Wraparound Milwaukee was stored in Synthesis, a database managed by Wraparound Milwaukee. Risk assessment data were stored in a separate database.

DCDS provides all juvenile justice services in Milwaukee County, with approximately 40 percent of juvenile justice youth enrolled in Wraparound Milwaukee. For this reason, the Milwaukee County implementation team included administrators, quality assurance staff, and IT workers from both agencies. Additionally, the implementation team included juvenile judges and probation supervisors to promote SPEP™ buy-in.

JJRRI's Approach to SPEP™ Implementation

Resources

As a function of participation in the OJJDP initiative, JJRRI demonstration sites had funding for staff time, particular for a JJRRI program manager, which was a critical resource for implementing the SPEP™. The program manager led the work of the implementation team and served a critical organizing function.

In addition, through JJRRI, extensive TA was available from the SPEP™ developers at Vanderbilt University and from the Georgetown Center for Juvenile Justice Reform. With these TA resources, the general JJRRI approach was that the TA providers first attempted to "install" the SPEP™ in the site during a first cycle of SPEP™ ratings and program improvement. During the first cycle, which is discussed in this report, the SPEP™ rating process was largely directed by the TA team from Vanderbilt. JJRRI's strategy was to use the first cycle to train staff and put in place the necessary implementation drivers so that the sites could then implement the SPEP™ themselves in future cycles.

Implementation Sequence

JJRRI approached all SPEP™ elements in parallel, attempting to obtain data on all four SPEP™ elements in the first round of ratings, followed by program improvement on the elements that required the most improvement.[6] When some elements proved difficult to implement, notably assessing the risk distribution of program clients, some JJRRI sites turned to producing interim reports to program providers to provide feedback on the SPEP™ elements that were available. In addition, even though JJRRI sites already had considerable stakeholder support, which was a key feature of sites' JJRRI grant applications, TA providers began early to buttress the support of the many stakeholders needed for all aspects of the SPEP™-guided CQI process, including program improvement.

JJRRI's TA providers approached SPEP™ implementation with a planned and sequential approach. Implementation was adaptive to each site's progress and challenges, so that while challenges in one area were being addressed, implementation often proceeded on complementary areas. The planned implementation sequence was approximately as follows:

- **Planning:** Diagnostic assessment of available site data; stakeholder engagement and introduction of the SPEP™ rating system; and organization of a local implementation team led by a program manager and made up of representatives throughout the juvenile justice system, such as from juvenile court, probation, service providers, and research.
- **Launching the SPEP™:** Identifying the programs to be rated by the SPEP™ through an assessment of the array of programs, classifying therapeutic programs according the SPEP™ service type, and constructing a rating system for a site's quality of service-rating.
- **Conducting a first round of SPEP™ ratings:** Conducting pilot SPEP™ ratings, making needed improvements to data systems and data-generating systems, and completing a first round of SPEP™ ratings.
- **Communicating results and planning program improvement:** Communicating initial SPEP™ results with programs and other key stakeholders, identifying areas of program improvement, and providing TA to help guide program improvement.

Although SPEP™ implementation roughly followed this sequence, implementation was adapted to the circumstances of each site. Some JJRRI sites proceeded to rate the set of community-based services before beginning to rate residential services.

The TA providers facilitated monthly calls with the implementation team at each site over the course of the grant to discuss implementation progress and challenges. TA providers also facilitated biannual cross-site calls, focused on topics that were relevant to the tasks currently being undertaken by demonstration sites (e.g., risk assessments, stakeholder buy-in and communication, and program improvement). One all-site meeting was hosted in spring 2015 with presentations by the demonstration sites, TA providers, OJP, and jurisdictions with past SPEP™ experience.

Planning

Initial engagement with JJRRI demonstration sites began late in 2012 with a series of phone calls, diagnostic worksheets, and gap analysis reports to facilitate the initial exchange of information among sites, TA providers, and the evaluation team. Two technical worksheets were completed by each site: a SPEP™ diagnostic worksheet and a cost diagnostic worksheet. The SPEP™ diagnostic worksheet focused on the availability of data elements required to produce SPEP™ ratings. This worksheet introduced sites to SPEP™ data elements and required an initial investigation into the availability of the

data and processes by which data may be collected. The cost diagnostic worksheet collected information on the contracting arrangements between juvenile justice agencies and program providers. This worksheet also focused on the data that would be required to assess the impact of the SPEP™ on changes in program costs.

Each site completed a "gap analysis," which focused on the larger juvenile justice context and collected a broad range of information on the local juvenile justice context. This information included local stakeholder support for the SPEP™ (including support from stakeholders in the legislature and courts, juvenile justice staff, and providers), organization culture (including how evidence-based practices were perceived), and juvenile justice system processes and procedures (including decisionmaking, the use of risk assessments, and the number and type of programs that serve juvenile justice youth).

SPEP™ implementation teams were organized at each JJRRI site, and consisted of approximately 20 members led by a program manager, Implementation teams varied across the sites but typically included representatives throughout the juvenile justice system, including juvenile court, probation and community corrections, service providers, and research and information systems divisions.

In spring 2013, implementation began with a stakeholder engagement meeting at each demonstration site. TA providers, the evaluation team, the SPEP™ program manager, and local stakeholders gave presentations at these meetings. TA providers coordinated planning with demonstration sites to discuss communication strategies and timing for stakeholders. Particular importance was paid to ensuring representation from the broad set of stakeholders whose buy-in was required for SPEP™ implementation and program improvement, including representatives of the community, court, juvenile corrections, and local policy advocates.

This was followed by a smaller meeting with the local implementation team, focused on developing an individualized implementation plan for the site. TA providers facilitated small group discussions around SPEP™ tasks and staffing requirements. For tasks that required more substantial and prolonged efforts, such as risk assessment analysis, data system enhancements, and SPEP™ sustainability, working groups were formed to support the work. At the end of this meeting, implementation teams were positioned to begin SPEP™ data collection.

Launching the SPEP™

Implementation began by assessing the array of juvenile justice programs to be included in the SPEP™, classifying programs into therapeutic services types, and establishing a quality of service rating scheme—all while continuing to build stakeholder support.

ASSESSING PROGRAM ARRAY

The first step was to identify juvenile justice programs that could be rated with the SPEP™ by collecting service descriptions and information on data availability from programs serving juvenile justice youth. The SPEP™ only rates services that take a therapeutic approach (versus a "control" approach) to juvenile justice youth, based on Lipsey's meta-analytic findings that therapeutic services are the ones with the potential to reduce recidivism. Thus, this step involved identifying therapeutic services.

This work varied significantly between sites depending on the size of the site and the availability of program information kept in MIS and other data systems. For example, identifying all juvenile justice programs across a state may be much more labor intensive than identifying juvenile justice programs within a county or city.

Although this task seems fairly straightforward, it sometimes required considerable coordination across agencies in JJRRI sites. In Iowa, for example, community-based program information needed to be collected across three judicial districts, and residential program information needed to be collected at the state level. To assess program information and data availability across districts, Iowa considered programs that provided services to youth in only one district or in multiple districts. Information on service descriptions, data availability, and funding streams was collected from juvenile justice programs, program provider contracts, and other agencies including DHS. In cases when program information was not available, site visits were conducted to discuss services provided by each program. In contrast, Delaware was well positioned to complete this task efficiently as its data reporting system held considerable information about the programs that were being rated on the SPEP™. Because all contracts ran through DCSD, this enabled data assembly regarding service provision.

At this early stage in SPEP™ implementation, TA was provided to the implementation team to educate members on SPEP™ requirements for ratings and how to review program profiles at each site to ensure they met the criteria for inclusion in the SPEP™. One criterion from the SPEP™ developers that required considerable attention was that each service's SPEP™ rating must be based on a cohort of at least 10 youths served who had risk assessments completed within the six months before entry into the service.

CLASSIFICATION OF SERVICES

Once the programs to be rated on the SPEP™ were identified, programs were matched to one of the therapeutic service types identified by Lipsey's research. In JJRRI, this began with on-site training from TA providers to discuss the classification process and how to identify distinct services in a program, which may require "unpacking" service components to determine primary and supplemental service profiles of a program (e.g., determining the main focus of the program versus other distinct service components).

Two special populations that generated considerable interest were juvenile substance abusers and juvenile sex offenders. However, the SPEP™ does not currently have separate ratings of effectiveness for specialized populations. Therefore, although general SPEP™ ratings could be applied, JJRRI sites were unable to apply more specific ratings to services that address these populations and needs.[7]

The amount of work involved in correctly classifying services largely depended on how much program information was contained within data systems and whether staff were required to complete site visits to collect classification data. If adequate program information did not exist in MIS or contracts, or if program information was out of date or inaccurate, the task of classification could prove very time consuming to coordinate with programs to access service information.

In Delaware and Iowa, much of the program information necessary to complete classification was contained in program contracts, so they were able to classify the programs that they wanted to rate within three months. In contrast, Milwaukee County attempted to rate a considerably larger number of programs, but program information was not stored in data systems or contracts. SPEP™ staff were, therefore, required to conduct site visits with more than two dozen agencies to collect service information in order to correctly classify programs. As a result, classification took more than eight months to complete.

When one program delivers multiple therapeutic services, each relevant service must be rated separately. Before that can be done, those distinct therapeutic services must be identified in a process of "unbundling." In Delaware, only community-based services were rated during the first round of SPEP™ ratings, which was much more straightforward than rating services provided in residential programs. Staff and TA providers began examining residential programs to identify and "unbundle" therapeutic services that could be rated on the SPEP™ in a second round of ratings.

RATING THE QUALITY OF SERVICE DELIVERY

In the SPEP™, the quality of service delivery concerns whether programs have processes to deliver services consistently and with fidelity. In this context, "quality" is similar to its usage in "quality assurance."[8] Quality-rating schemes were tailored to the sites and were more site-specific than other SPEP™ elements. TA providers helped the sites develop tailored rating schemes based on existing quality measures and establish appropriate cut-offs to distinguish between high-, medium-, and low-quality services.

In JJRRI, the quality-of-service ratings rely on whether there are written service protocols or manuals; whether staff are trained according to that protocol and receive ongoing staff training; whether there are procedures in place to monitor adherence to protocols and consistent service delivery; and whether there are procedures to take corrective actions in response to departures from the service model. The sites piloted their rating scheme and shared the results with the TA providers prior to moving forward with larger-scale ratings of the quality of service delivery.

Data collection to generate quality-of-service ratings can be a relatively painless process for sites that have quality monitoring or audits built in to contracts or other programmatic quality assurance requirements. In Iowa, for example, quality measures for residential providers were included in Requests For Proposals from service providers. However, for sites without a manualized quality assurance process, site visits were necessary to understand programs operations.

Considerable communication with program and agency stakeholders may be necessary to gain access to quality-of-service records because this process may uncover deficiencies. JJRRI sites typically worked with TA providers to draft written and oral communication about the SPEP™ to educate and maintain or increase buy-in, while attempting to allay concerns that the process would be used to penalize programs. In JJRRI, TA providers facilitated on-site trainings on effective communication strategies with programs. Semiannual site visits also provided opportunities for TA providers to meet with program providers to discuss the SPEP™ rating and the importance of assessing program effectiveness. Meetings with juvenile justice program stakeholders provided a forum for staff to discuss challenges and barriers to SPEP™ data collection requirements, including frustration with data requests that increase the work burden for service providers, and misunderstandings over the types of data that need to be routinely collected and/or how data should be reported to the SPEP™ implementation team.

Conducting a First Round of SPEP™ Ratings

A pilot rating phase, with a small number of services, took at least six months to complete in JJRRI sites. This was due in large part to the issues encountered with the individual client-level data concerning dosage and risk needed to conduct the ratings.

When sites have automated data already available, then the initial round of SPEP™ ratings can be retrospective. That is, the program-level data can be collected about a cohort of youth who participated in services during a time period that has just ended. In contrast, when data systems need to be established, then initial SPEP™ ratings must be done prospectively, with youth who have not yet received services. Once the system for collecting data is established, data for a 12-month service will take another year to accumulate. But the systems may also take some time to establish, further delaying the first round of ratings. Especially when the first round takes an extended period of time to complete, changes (i.e., improvements) to programs may be made even before the first round of ratings, especially once program stakeholders understand the underlying reasons why the SPEP™ elements improve program effectiveness. In this situation, the first round of ratings may not represent a true baseline state. JJRRI sites varied in this regard.

AMOUNT OF SERVICE

SPEP™ ratings require detailed data on the amount or dosage of services provided to each client. This includes data on service duration and the number of face-to-face contact hours. There are several ways in which sites can collect dosage data. Contracting mechanisms may require programs to regularly submit the amount of services provided to youth. If dosage is not regularly submitted, coordination with programs may uncover program-specific dosage recording systems. When detailed dosage information is not directly available, dosage may sometimes be calculated from information about the contracted length of service and the number of service contacts made weekly or monthly.

Unless all juvenile justice program contracts at a site require the regular submission of dosage data, sites will likely have to independently collect some form of dosage data. All three JJRRI sites had some fee-for-service contracts in place that allowed for easy dosage data collection with many programs. For other programs, dosage data was typically requested and/or collected in parallel with the program-level data collection tasks of classification and quality-of-service ratings.

One particular complication arises around the rating of dosage regarding name-brand programs. SPEP™ dosage requirements are generally based on the evidence for a type of service (e.g., family therapy) based on the meta-analyses of all programs of that type, including both name-brand and

generic programs. But some name-brand programs, such as Functional Family Therapy and Multisystemic Therapy, have set their own standards and accumulated sufficient evidence of effectiveness based on those targets to warrant their own dosage criteria in the SPEP™ rating process. This arose in one JJRRI site, and SPEP™ developers worked with the site to tailor the SPEP™ rating scheme.

RISK OF REOFFENDING

Risk assessment data is the final element required to calculate SPEP™ ratings. Risk data from the most recent risk assessment completed before initiation of a service, within the preceding six months, are collected for all youth who participated in a service. Ideally, the data are derived from a juvenile risk assessment that has been locally validated to ensure that the assessment correctly discriminates between high-, medium-, and low-risk youth. It is also important that the assessment is completed consistently.

In theory, risk assessment data can be easily integrated to produce SPEP™ ratings. Case management software allows users to enter, manage, and store assessment results and reports and typically allows them to track clients and assessment information over time. Risk assessment tools are also typically designed to be conducted on a regular schedule (e.g., every six months), theoretically allowing SPEP™ sites the ability to pull data from the most recent assessment for the youth in SPEP™ programs.

However, JJRRI sites encountered significant challenges with their risk assessment data, even though all three demonstration sites had already implemented the use of risk assessments. TA providers spent considerable time working with sites to adapt risk assessments to the SPEP™.

Delaware. In 2012, Delaware began using the Positive Achievement Change Tool (PACT) as its risk and needs assessment tool, following adjudication, only for youth who were disposed to community supervision. Thus, the risk data could not inform judicial placement decisions and were not available for youth in residential settings. Because risk data are a critical requirement for the SPEP™ ratings, only community-based programs contracted through DYRS were rated on the SPEP™.

On the SPEP™ rating, a service generally receives points for the proportion of clients who exceed the threshold for high risk. However, the PACT rating system did not initially classify youth into high/medium/low categories of risk. As a result, TA providers worked closely with the implementation team to help establish appropriate cut-off points.

Iowa. In 2007, Iowa began to use the Iowa Delinquency Assessment (IDA) to assess the risk and needs of juvenile justice youth. A short-form IDA was completed with all youth at intake to assess their level of risk and eligibility for diversion. A long-form IDA was completed with youth who rated in the moderate- to high-risk area on the short form and used to assess service needs and placement before adjudication. Risk information was stored in the Iowa Court Information System, a statewide data system that stores juvenile and criminal justice case information and is regularly updated by juvenile court staff.

A key issue was that the quantitative risk assessment data (short-form IDA) was often older than the six-month inclusion criteria for the SPEP™. Although some qualitative data were available from the long-form IDAs, this data could not easily be transformed into the quantitative risk data needed. The TA providers worked with the site to devise proxy risk measures from the long-form data and/or available criminal history information for the preliminary round of SPEP™ ratings. In addition, a considerable effort was put in place to establish consistent risk assessment so that the data would be more readily available in future cycles of SPEP™ ratings.

Milwaukee County. In 2012, DCSD began using the Youth Assessment and Screening Instrument (YASI) as its risk and needs assessment to inform recommendations and decisionmaking about juvenile supervision and services. At the start of Milwaukee's JJRRI grant, human service workers (essentially, juvenile probation officers) completed risk assessments with youth. However, by the time JJRRI began in 2013, risk assessments were not being completed systematically. Judges also did not consistently allow risk assessments to be introduced in juvenile court hearings or used to guide placements.

Therefore, stakeholder support for risk assessment was lacking, both from the staff who conduct risk assessments and from key users of the risk assessment data. Considerable TA was involved in working on these issues. To increase familiarity with and understanding of the YASI as an assessment tool to guide decisionmaking, DCDS also implemented training (and boosters) on using the risk assessment (and on motivational interviewing techniques) for human service workers, Wraparound Milwaukee staff, and members of the judiciary. In addition, a policy was implemented requiring the risk assessment to be completed with every new youth referred to DCDS for a delinquent act.

Finally, at the beginning of JJRRI, Milwaukee lacked an MIS that allowed for easy integration of risk assessments. The risk assessment data needed for the SPEP™ had to be entered by hand. The site put considerable effort into transforming management information systems so that they would easily accommodate the risk data for future cycles of SPEP™ ratings.

In sum, across the JJRRI sites, the SPEP™ risk element involved each of the SPEP™ implementation drivers discussed earlier. It often involved improvements to data systems and data-generating systems,

as well as considerable work to buttress stakeholder support. Each of these, in turn, involved considerable TA support.

It is important to note that in JJRRI, the effort to improve risk assessment and risk data was also motivated by another juvenile justice reform. As mentioned in the beginning of this report, JJRRI worked to implement dispositional matrices as a tool for dispositional decisionmakers. The central idea behind such dispositional matrices is that dispositional decisions should be aimed at reducing reoffending and therefore should be guided largely by delinquency risk (Boglivio, Greenwald, and Russell 2015), with commitments and other expensive and restrictive services reserved for those youth with the highest risk of reoffending. Dispositional matrices recommend dispositional options on based risk assessment information combined with other information related to the current case (e.g., most serious charge). Because the SPEP™ and dispositional matrices share a concern with reliable risk assessment, improving risk assessment was a major objective of work at all JJRRI sites.

SPEP™ RATING PROGRESS

Two sites struggled to produce a pilot and first round of SPEP™ ratings until well after the first year of implementation. The types and number of programs that were included in the SPEP™ ratings varied considerably among sites. The sites varied in the data elements that posed challenges, but all sites were challenged by the risk assessment data.

Delaware. Delaware rated only community-based programs because risk assessments were not being completed with youth placed in residential programs. Delaware also limited its SPEP™ ratings to the community-based services that had most recently been implemented in Delaware and not yet been assessed through any formal procedures. Six community-based services provided through one organization were rated, separately, in each of three counties, for 18 possible ratings. Several of these services did not have enough clientele during the cohort period to generate SPEP™ scores. Thirteen services received full SPEP™ rating, and one received a "provisional" rating. Data were available to allow the first cohort to be rated retrospectively, and the SPEP™ reports were completed within the first year of implementation.

Iowa. In Iowa, where data systems and stakeholder support were particularly strong, the pilot phase generated full-rating SPEP™ ratings for four community-based services in one community program and 22 services in one residential program. First round SPEP™ ratings were then completed for 15 additional community-based services across three community programs and 30 additional services across three residential providers. In sum, the first round of SPEP™ ratings described 19 community-based and 52 residential-based services.

However, all of Iowa's based SPEP™ ratings of residential services, and about half of the ratings of the community-based services, were classified as either "preliminary" or "provisional" due to limited or out-of-date risk assessment data. SPEP™ reports, including program improvement plans, were reviewed with the first round of rated programs in early 2015, approximately 19 months after the SPEP™ was launched. Although Iowa was able to generate SPEP™ ratings for more programs than the other JJRRI sites by the end of the evaluation, programs in the effort included only those that serviced the greatest number of youth, including programs in urban locations and residential facilities.

In addition, one Iowa community-based service provider providing two substance abuse services believed that their client-confidentiality provisions did not allow them to share individual-level data for past services and clients without informed consent. They then modified their intake and client informed consent forms and established data agreements with CJJP and the Department of Public Health to allow this data-sharing prospectively, but they were unable to participate in the first round of SPEP™ ratings.

Milwaukee County. In Milwaukee, more than 200 juvenile justice programs were classified. However, because of the inconsistent use of juvenile risk assessments and weak data-generating systems, only six community-based services across four providers and five residential services across two residential providers had received SPEP™ ratings by the end of 2015. Three community-based programs received full-rating reports, and three received "provisional" rating reports. Only one residential program received a full-rating report.

Communicating SPEP™ Results and Planning Program Improvement

SPEP™ ratings indicate to what degree services are consistent with the evidence of what reduces recidivism. They also indicate where there is room for a service to improve. If a program is low on dosage, there is the potential to increase the number of contact hours and/or duration. If a program is not rated highly on risk, there is the potential to serve more high-risk youth. However, providers are often particularly concerned about being rated on elements beyond their control. Programs may have a limited capacity to increase the dosage of services provided, if billable contact hours are limited by contract requirements or judicial decisions. In addition, programs typically have little control over the risk level of the clients referred to them.

From the beginning of JJRRI, the TA providers worked with the implementation team and stakeholders to reinforce the message that the SPEP™ ratings are a tool for improving program

effectiveness rather than an audit or an attempt to identify and sanction ineffective programs. Once the SPEP™ ratings were complete, the program managers and TA providers worked to ensure that the reports and communication with service providers also communicated that message.

The TA providers often stressed that jurisdictions need an array of different types of services. For example, in the SPEP™ rating scheme, CBT is the service type with the highest potential to reduce recidivism, but this does not imply that all services should be replaced with CBT. Therefore, the SPEP™ developers have also developed POP ratings, which compare each service to the highest potential SPEP™ rating for that service type, under the assumption that the service cannot change its service type.

In JJRRI, even when risk data were found inadequate to generate full SPEP™ ratings, it was deemed important to provide preliminary feedback to those programs that were being rated on the SPEP™, even before the systems, policies, and practices of risk assessment could be improved to meet SPEP™-rating requirements. Therefore, TA providers worked with JJRRI sites to adapt SPEP™ reporting mechanisms to allow the reporting of "provisional" and "preliminary" SPEP™ ratings—with appropriate caveats and messaging—that used the incomplete or proxy data that were available. For example, for some Iowa programs, criminal history data were able to provide a proxy for missing risk assessment data. In addition, the site program managers, in consultation with the TA providers, worked to provide partial feedback to services concerning the other elements rated on the SPEP™.

PROGRAM IMPROVEMENT PLANNING

Initial SPEP™ ratings often find considerable room for program improvement, with POP ratings (which take service type as fixed) often below 50 percent. In some CQI efforts, program improvement is left to the programs themselves. In JJRRI, however, sites were interested in more TA for developing plans and processes to assist their use of SPEP™ ratings to guide program improvement.

In JJRRI, TA providers made site visits to each demonstration site to work with the implementation team on a strategy and process for program improvement. Program improvement work groups were established and met regularly to draft written SPEP™ reports for providers, discuss potential challenges and barriers to program improvements, and develop processes to guide program improvement. Some of the changes needed to improve program effectiveness required complex, multistep efforts. The TA providers generally recommended a process called the Plan-Do-Study-Act,[9] which involves a rapid cycle of attempts to identify solutions, try them, and see how they are faring.

Each of the four SPEP™ elements has different program improvement considerations. Service type would not seem to be amenable to change. However, some service types receive additional rating points when secondary services are included that have been found to raise their effectiveness. Program improvement considerations, therefore, may include trying to add such secondary service components.

The SPEP™ quality element is typically a focus of program improvement efforts, including training and certification efforts, quality assurance and monitoring, and processes for correcting drift in fidelity to the service model. JJRRI reports have often found that the quality ratings are those with the most potential for a program to improve its effectiveness; this was typical in JJRRI sites as well.

Although risk data features heavily in implementation challenges for the SPEP™ ratings, risk featured much less heavily in program improvement plans because the distribution of risk among clients referred to programs is generally not something that programs can control and may not be part of their program improvement efforts. However, that there are times when programs have a significant role in the risk level of cases accepted or retained. Some service contracts are not specific to juvenile justice youth, such as when they are contracted or funded through mental health or child welfare agencies. Such service providers sometimes refuse delinquent youth, especially high-risk youth, or terminate them early. For services with mixed clientele from different referral sources, the SPEP™ rating should be based only on delinquency cases. Program exclusion of high-risk youth would lead to low ratings on the SPEP™ risk element, and early termination of high-risk youth would lead to low ratings on the SPEP™ dosage element. In such cases, program improvement might involve changing practices to effectively serve higher-risk youth.

Finally, there is often room to improve dosage. Understanding the reasons why youth are not receiving recommended levels of service is key to program improvement. In JJRRI, some services were contracted through agencies that primarily serve other youth (e.g., mental health and/or child welfare referrals). Those agencies may have targets for service hours and/or service duration that differ from the dosage targets in the SPEP™. Recall that the SPEP™ targets are based on studies of effectiveness specifically in reducing reoffending. But mental health funded services may base their targets on considerations other than recidivism reduction. Program improvement in such situations may involve complex negotiations across agencies and programs.

PROGRAM IMPROVEMENT TIMING

When used in a CQI process, the SPEP™ is embedded in a regular cycle of SPEP™ ratings, program improvement, and re-ratings. Typically, the cohorts of service clients used to generate SPEP™ ratings (the individual-level elements) are 12-month service-entry cohorts. A typical period for program

improvement before the next rating cycle is also 12 months. Such a schedule then requires three years for two rounds of ratings, with a year for program improvement. If data for a first round of rating can be collected retrospectively, based on youth recently served by a program, then that first cycle may be shortened. But for the SPEP™ to guide program improvement, program improvement planning can begin to be developed only after reports of the SPEP™ ratings have been shared with providers.

When the rating process is extended, as is typical in a first round of ratings, the first CQI cycle is often affected. Once the basis of the SPEP™ is understood, providers may take preliminary steps to improve their programs' effectiveness at reducing recidivism, such as improving their quality assurance processes, even before the first round of SPEP™ ratings is completed. In these cases, the first round of SPEP™ scores may not reflect a true baseline state because preliminary improvements have already been incorporated into the initial ratings.

6. Conclusion and Lessons Learned

This report provides a window into implementation of the SPEP™ through the experiences of the JJRRI sites over three years. Although a key goal of SPEP™ ratings is to guide program improvement, this report is limited to the implementation of the critical first round of ratings and its requirements. Chapter two described four critical drivers of SPEP™ ratings, namely systems for data and data generation, stakeholder support, implementation teams, and technical assistance. Chapter four described implementation of the first round of SPEP™ ratings at the JJRRI sites, from which we draw several basic lessons:

- Although the SPEP™ may seem technically straightforward as a rating system for programs serving juvenile justice–involved youth, its implementation may take considerable time and effort. An extended implementation period is generally required to accomplish the first round of ratings and to put in place the necessary systems for the next round. As a result, the planned three-year project periods for JJRRI (which have been extended) did not generally allow for completion of program improvement.
- Support from many juvenile justice stakeholders is critical for the SPEP™ rating system to be used effectively to improve the effectiveness of services provided to juvenile justice youth.
- Working on how the SPEP™ results will be communicated and used from the initiation of the process is critical if the SPEP™ ratings are to deliver on their promise of improving program effectiveness.
- Considerable TA is needed for technical aspects of the SPEP™ and its required data. TA is also needed to build and buttress stakeholder support and to use SPEP™ ratings for program improvement.

The SPEP™ as a Quality Rating and Improvement System

The SPEP™ is but one of several quality rating and improvement systems (QRIS) being used to improve the implementation and effectiveness of government-supported programs. For example, the Administration for Children and Families within the US Department of Health and Human Services supported a rating system for child care, in which 26 jurisdictions participated (Tout et al. 2010). Implementation of these rating systems generally takes two to fo0ur years (Paulsell, Tout, and Maxwell

2013). A study of implementation of the child care rating system in five pioneer states reached conclusions similar to the current report, with a stress on the need for resources, political support, and pilot work (Zellman and Perlman 2008).

Similarly, for application to Head Start programs, Derrick-Mills and colleagues (2014) reviewed literature on data use for CQI from education, health care, nonprofit management, public management, and organizational development, with a focus on within-organization change. Paralleling our discussion of a broad range of stakeholder support, they stress the significance of leadership support that is inclusive and participatory and the importance of the broader context in which the program operates. Paralleling our discussion of data, systems, and TA, Derrick-Mills and colleagues (2014) highlight the importance of data and analytic capacity. They also stress the importance of time and resources, which were critical ingredients in JJRRI provided through grant funding for staff time and TA provision.

In short, reports on implementation with other QRIS in other domains reach similar conclusions about the complexity of the undertaking and implementation challenges.

Outcomes and Performance Standards

The SPEP™ does differ in one important way from other such rating systems. Many systems use performance measures that are developed based on the consensus of experts (Derrick-Mills et al. 2014), and local input into those rating systems is an important vehicle for securing stakeholder support. The performance standards may be an important step for program improvement, but typically they are not based on strong prior evidence that the criteria in the rating system are related to outcomes. In contrast, because the SPEP™ is based on considerable prior research, it can be thought of as a tool for bringing the research on effectiveness in reducing recidivism to bear directly on practice. This also means that the SPEP™ is much more standardized across jurisdictions than rating systems that are developed locally, even though the SPEP™ developers have worked with jurisdictions to tailor the system somewhat to local data.

Consider how the SPEP™ evidence base relates to the standards for dosage. When locally developed as performance standards, criteria for dosage will often be a consensus standard based on practicality considerations, such as available funding, current practice, and some sense from practitioners and service providers about program effectiveness. Yet service providers typically have little access to data on participant recidivism to ground their sense of when they are effective. Especially if developed in tight fiscal circumstances, the resulting consensus criteria might very well be

too low to actually produce the hoped-for reductions in recidivism for the high-risk youth most in need of services. In contrast, because the SPEP™ dosage criteria are based on prior research, rather than being the output of a negotiation process, the SPEP™ dosage standards (and the SPEP™ ratings) serve as inputs to a system improvement process that may involve negotiations over funding for effective dosage of services.

That the SPEP™ rating is standardized means that its implementation timeline is different than implementing a QRIS, in which developing performance-based standards is a first step of implementation. This obviously eliminates an important first step and would seem to shorten the expected implementation time. However, this is offset by other SPEP™ features. Because the SPEP™ elements are predetermined, the SPEP™ is largely resistant to the elimination of elements that will be slow to implement and that may require changes to data systems (although quality ratings have sometimes been skipped, see Redpath and Brandner 2010).

That the research base for the SPEP™ does connect the SPEP™ criteria to outcomes is a major selling point of the effort; it provides a reason for the system reforms that may be necessary to implement the SPEP™. At the same time, because the SPEP™ standards themselves are not amenable to much local input, that tool for building stakeholder support is not available. Therefore, it is all the more important to intentionally build and buttress stakeholder support from the beginning of the SPEP™-rating process.

The SPEP™ as a Vehicle for Juvenile Justice Reform

Although implementation of the first round of SPEP™ ratings may take considerably longer than the rating scheme's simplicity might suggest, the SPEP™ seems to be a quite effective vehicle for juvenile justice reform and for putting a jurisdiction on a much stronger platform for evidence-based decisionmaking. The very effort to implement the SPEP™ tends to expose some system weaknesses and point the way to addressing them. Although jurisdictions often have some of the necessary infrastructure in place to assess some of the SPEP™ dimensions, the process of trying to assess programs with the SPEP™ may reveal that youth risk assessment, data systems, monitoring of the amount of service delivered, and/or quality assurance processes warrant improvement in important ways.

Equally important, because the SPEP™ rating is built on a strong evidence base of program effectiveness in reducing recidivism, the SPEP™ rating system provides an effective vehicle for

translating research into practice, which affords a strong rationale for improving these systems. That is, improving these systems in the context of SPEP™ implementation is driven by an evidence-based promise of recidivism reduction. This linkage to outcomes may have been lacking previously.

At the program level, the SPEP™ is based on research that has found that some types of services are much more effective than others in reducing recidivism—if delivered consistently and with fidelity. Previously, a system may not have had much reason to describe the key elements of programs in a systematic way, without compelling evidence about which types of programs reduce recidivism more effectively. In addition, there may have been little reason to assess whether programs had procedures to ensure fidelity to any model, in the absence of evidence of which models were effective.

This is similar for the service dosage that is delivered to each youth. The research base that underlies the SPEP™ indicates what dosage of services was found effective in reducing recidivism; with lower doses, the same benefit cannot be expected. Attempting to compile SPEP™ ratings sometimes reveals that jurisdictions have not built systems that systematically record the contact hours received by each youth for each service, and there may not have been sufficient reason to build such systems previously. That is, without evidence to suggest how much of a service is needed to reduce recidivism, it may have been reasonable to leave the amount of service to providers' discretion.

Finally, the SPEP™ ratings turn critically on risk assessment data. Reliable and validated risk assessment is important for effective dispositional practice; it plays a key role in many juvenile justice reform efforts. The research on which the SPEP™ is based, and the potential of the SPEP™ to help translate that research to practice, provides another strong reason to improve risk assessment systems and practices. In JJRRI, risk has had a central place in efforts to improve both dispositional practices and service effectiveness, and, in turn, to help jurisdictions more effectively use their service dollars.

When some systems needed for the SPEP™ are missing, the SPEP™ can motivate the building of a data and reporting platform, which, in turn, improves a jurisdiction's capacity for self-assessment and self-monitoring. Decisionmakers then have much better descriptions of what services are being provided to which youth, and enhanced ability to describe and monitor the services that are being provided.

In sum, for jurisdictions in which the systems are already in place, to allow SPEP™ assessment to proceed expeditiously, the SPEP™ can fairly quickly become a vehicle for program improvement. But for many jurisdictions, implementing the SPEP™ for the first time proves harder than anticipated because it first entails improving assessment, reporting, or data systems. In such jurisdictions, the program improvement from the SPEP™ may be delayed, but the SPEP™ may have the important added benefit of

building the general capacity for self-assessment and monitoring that is a key to evidence-based juvenile justice systems reform.

Appendix A. JJRRI Sites

Delaware

Site description	The state of Delaware is a unified court system comprised of Sussex, New Castle, and Kent counties. Delaware is one of the least populated states in the United States, with approximately 900,000 residents, 13 percent of whom are between the ages of 10–19.
Lead agency	Department of Youth and Rehabilitative Services (DYRS)
Lead agency description	DYRS is one of four divisions within the Department of Services for Children, Youth and their Families (DSCYF). The DYRS Community Services Unit provides community supervision to approximately 3,000 preadjudicated and adjudicated youth throughout the state. DYRS oversees contracts with providers offering residential and nonresidential services.
Goals	Reduce crime and delinquency; focus more heavily on youth prosocial skill development; provide a responsive supervision continuum of services based on the risk of reoffense; develop recidivism measures for community-based services.
Eligible juvenile justice programs	Community-based programs implemented within one year before SPEP™ implementation.
Source of funds for eligible juvenile justice programs	Community-based services are provided by private and nonprofit agencies; juvenile justice services are funded by DYRS.
Data systems and management at the time of SPEP™ implementation	DSCYF maintains the Family and Child Tracking System, a statewide information system that provides real-time case tracking across DYRS, the Division of Family Services, the Division of Prevention and Behavior Health, and the Division of Management Support Services. The system contains more than 30 categories of datasets, including demographic information, assessments, case planning, legal status, disciplinary, service utilization, and placement authorization. DSCYF staff also has access to the Delaware Criminal Justice Information System, which maintains case-level data for court cases.
Risk assessment at the time of SPEP™ implementation	The Positive Achievement Change Tool (PACT) was implemented statewide in 2012. PACT assessments are completed by DYRS workers with youth who receive community service and placement. PACT is only administered following adjudication and therefore has no bearing on judicial decisionmaking.
Number of participating community-based services in first round SPEP™	13 services within four community-based programs.
Number of participating residential services in first round SPEP™	0
Number of youth in community-based services in first round SPEP™	338
Number of youth in residential services in first round SPEP™	0

Iowa

Site description	Iowa is a unified court system comprised of eight judicial districts, which each encompass five or more counties. Each district is autonomous, without standardized community-based services. Each district has one juvenile court that presides over Child in Need of Assistance, adoption, delinquency, and commitment cases for youth between the ages of 10 and 18. Juvenile courts are supervised by a chief juvenile court officer who oversees juvenile court operations, probation, and case management, as well as local research and program development.
Lead agency	Iowa Division of Criminal and Juvenile Justice Planning (CJJP)
Lead agency description	CJJP is located within the Iowa Department of Human Rights and serves as the Iowa Statistical Analysis Center. CJJP administers federal and state grant programs to fund local and state projects to prevent juvenile crime, provide services to juvenile offenders, and otherwise improve Iowa's juvenile justice system. CJJP carries out research, policy analysis, program development, and data analysis activities to assist policymakers, justice system agencies, and others to identify issues of concern and improve the operation and effectiveness of Iowa's justice system. CJJP staff provide a justice system information clearinghouse service to system officials and the general public.
Goals	Ensure that juvenile justice sanctions and services are effective; ensure that youth receive appropriate sanctions and services based on their needs and level of risk; implement a comprehensive method to measure the effectiveness of juvenile justice programs that can be standardized across the state.
Eligible juvenile justice programs	Community-based programs in first, third, and sixth judicial districts and residential programs throughout the state.
Source of funds for eligible juvenile justice programs	CJJP and the Department of Human Services (DHS) fund district juvenile court offices, which control contracts with community-based services. Residential services receive funding directly from CJJP, DHS, and district court offices.
Data systems and management at the time of SPEP™ implementation	Iowa's judicial branch maintains the Iowa Court Information System (ICIS), a statewide information system comprised of juvenile and criminal justice processing information. ICIS is updated by juvenile court staff and contains juvenile risk information from the Iowa Delinquency Assessment (IDA). ICIS information is housed in the Justice Data Warehouse, maintained by CJJP.
Risk assessment at the time of SPEP™ implementation	The Iowa Delinquency Assessment (IDA) was implemented statewide in 2007.There are two forms of the IDA: a short form used at intake to determine the risk level of youth and whether they may be diverted from the court system or require more intensive supervision and service; a long form is used with youth who score moderate to high risk on the short form. The long form is used to determine appropriate services. The IDA is completed by juvenile court officers. The IDA was in the process of being validated at the time of QRIS installation.
Number of participating community-based services in first round SPEP™	19 services within nine community-based programs.
Number of participating residential services in first round SPEP™	52 services within four residential settings.

Number of youth in community-based services in first round SPEP™	501
Number of youth in residential services in first round SPEP™	411

Milwaukee County, Wisconsin

Site description	Milwaukee County is the largest county in Wisconsin, with a population of approximately 1 million; 25 percent of the population is under the age of 18. One juvenile justice court presides over all cases, with discretion to place youth in over 200 juvenile justice programs in the county.
Lead agency	Milwaukee County Delinquency and Court Services Division (DCSD)
Lead agency description	DCSD serves youth from the point of juvenile referral to the end of the disposition order. DCSD contracts with and monitors the administration of juvenile services, including the operation of a 120-bed juvenile facility.
Goals	Ensure that youth are systematically matched to programs based on risk and need assessments; assess the effectiveness of juvenile justice programs; engage in ongoing quality assurance and improvement.
Eligible juvenile justice programs	Community-based and residential programs in Milwaukee County.
Source of funds for eligible juvenile justice programs	DCSD funds approximately 40 community-based programs to provide services for juvenile justice youth. Wraparound Milwaukee[a] funds more than 200 services through support from DCSD, child welfare, Medicaid, and mental health resources.
Data systems and management at the time of SPEP™ implementation	DCSD maintains the Juvenile Information Management System, which tracks juvenile information from referral to the end of court-imposed supervision. Wraparound Milwaukee maintains Synthesis, a web-based client database that tracks referrals, services, care plans, case notes, and invoices for youth receiving services through their network of providers.
Risk assessment at the time of SPEP™ implementation	The Youth Assessment and Screening Instrument (YASI) was implemented in Milwaukee County in 2012, albeit on a limited basis. The YASI prescreen is used prior to referral to court to categorize youth into risk levels. If a youth scores in the low-risk category, he or she may be diverted from court; if the youth scores in the medium- to high-risk category, he or she will be assessed by the YASI full-screen with the intention to inform disposition. However, at the time of SPEP™ implementation, YASI was not yet being utilized to information judicial decisionmaking.
Number of participating community-based services in first round SPEP™	6 services within 4 community-based programs
Number of participating residential services in first round SPEP™	5 services within 2 residential settings
Number of youth in community-based services in first round SPEP™	105
Number of youth in residential services in first round SPEP™	91

[a] Wraparound Milwaukee provides mental health services to juvenile justice youth with a DSM-IV diagnosis. Wraparound Milwaukee provides case management services and contracts with approximately 175 programs throughout the county for direct service care.

Appendix B. Sample SPEP™ Scoring Sheet

Standardized Program Evaluation Protocol (SPEP™) for Services to Juvenile Offenders© Recalibrated version, 2013	Points Possible	Points Received
Primary and Supplemental Service Types [Identified based on definitions derived from the research]		
Primary Service Type for Program Being Rated Group 1 services (5 points) Group 2 services (10 points) Group 3 services (15 points) Group 4 services (25 points) Group 5 services (30 points)	30	
Supplemental Service Type Qualifying supplemental service utilized: Yes (5 points) No (0 points)	5	
Quality of Service Delivery [Determined from a systematic assessment of the relevant features of service implementation]		
Rated quality of services delivered: Low (5 points) Medium (10 points) High (20 points)	20	
Amount of Service [Determined from dosage 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 medium or high risk scores (greater than low): 0% (0 points) 75% (7 points) 30% (2 points) 85% (10 points) 50% (5 points) 95% (12 points) % of youth with high risk scores (greater than medium): 0% (0 points) 25% (8 points) 15% (3 points) 30% (10 points) 20% (5 points) 35% (13 points)	25	
Total SPEP™ Score	100	(Insert Score)

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

Appendix C. Sample Dispositional Matrix

Florida Department of Juvenile Justice Disposition Recommendation Matrix

(Staff should begin with the least restrictive setting within a particular disposition category. See Structured Decision-Making guidelines.)

Most Serious Presenting Offense	PACT Risk Level to Reoffend			
	Low Risk to Reoffend	Moderate Risk to Reoffend	Moderate-High Risk to Reoffend	High Risk to Reoffend
Civil Citation Eligible[1]	Level 1	Level 1	N/A	N/A
Minor[2]	Level 2 or 3a	Level 2 or 3a	Level 2 or 3a-c	Level 3a-c or 4
Serious[3]	Level 2 or 3a	Level 2 or 3a-b	Level 3a-c or 4	Level 3a-c or 4
Violent[4]	Level 2 or 3a-b	Level 2, 3a-c, or 4	Level 3a-c, 4, or 5	Level 3a-c, 4, or 5

[1] – Eligibility for civil citation is outlined in F.S. 985.12. Youth deemed ineligible for civil citation (based on community standards) should be reviewed under the "Minor" offense category based on the PACT risk level to reoffend.

[2] – All misdemeanor offenses.

[3] – Felony offenses that do not include violence.

[4] – Violent felony offenses (do not include misdemeanor assault and battery which are captured under "Minor").

Level 1 – Alternatives to Arrest
Level 2 – Diversion & Non-DJJ Probation
Level 3 – Community Supervision
Level 4 – Non-Secure Residential Commitment
(3a) – Probation Supervision
Level 5 – Secure Residential Commitment (High & Maximum Risk Programs)
(3b) – Probation Enhancement Services (ART, EPICS, LifeSkills, etc.)
(3c) – Day Treatment, MST, FFT, Minimum Risk Commitment

Updated January 2016

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

Notes

1. SPEP™ ratings range from 0 to 100. A sample SPEP™ rating tool is attached in appendix B.
2. For an in-depth review of the basis for the SPEP™, see Lipsey (2009) and Lipsey and colleagues (2010).
3. The literature has not found a level of risk so high as to render programs ineffective.
4. The SPEP™ considers risk of recidivism, but not risk of mental health issues, educational failure, or any other risk. In discussing risk and needs assessment, risk is often distinguished from needs, with risk referring to actuarial risk of recidivism and needs referring to issues that could be addressed by services, such as poor family functioning, mental health issues, susceptibility to deviant peers, educational deficiencies, and so on. Different risk and needs assessment systems rate domains somewhat differently (see Hoge, Vincent, and Guy 2012). There is considerable controversy in the research field concerning whether and how assessed needs should have combined with actuarial risk of recidivism (see Baird et al. 2013 and commentaries).
5. During the course of JJRRI, risk assessment was expanded to residential youth, making it possible to include residential services in later rounds of SPEP™ ratings.
6. An alternative approach would be to address the SPEP™ elements in a serial fashion, rating and attempting to improve some SPEP™ elements before the others, such as by rating and trying to improve the program-level aspects (i.e., service type and quality of service delivery) before assembling the data to rate SPEP™ client-level elements (i.e., dosage and risk). Although each SPEP™ element can be improved on its own, they achieve their full potential together. Thus, true SPEP™ ratings require all four elements.
7. SPEP™ developers have been working to develop such ratings, in light of a growing evidence base on these types of specialized programs, but such a rating scheme was not completed by the time of SPEP™ ratings under JJRRI.
8. This is quite distinct from quality of a delivered service, as assessed through direct observation by experts, which is not rated by the SPEP™.
9. See "Plan-Do-Study-Act (PDSA) Cycle," Agency for Healthcare Research and Quality, last modified April 10, 2013, https://innovations.ahrq.gov/qualitytools/plan-do-study-act-pdsa-cycle.

References

Andrews, Don A., James Bonta, and Robert D. Hoge. 1990. "Classification for Effective Rehabilitation: Rediscovering Psychology." *Criminal Justice and Behavior* 17 (1): 19–52.

Baird, Chris, Theresa Healy, Kristen Johnson, Andrea Bogie, Erin W. Dankert, and Chris Scharenbroch. 2013. *A Comparison of Risk Assessment Instruments in Juvenile Justice*. Rockville, MD: National Criminal Justice Reference Service. https://www.ncjrs.gov/pdffiles1/ojjdp/grants/244477.pdf.

Boglivio, Michael T., Mark A. Greenwald, and Mark Russell. 2015. "Assessing the Implications of a Structured Decision-Making Tool for Recidivism in a Statewide Analysis: Disposition Matrix for Court Recommendations Made by Juvenile Probation Officers." *Criminology and Public Policy* 14 (1): 5–49.

Derrick-Mills, Teresa, Heather Sandstrom, Sarah Pettijohn, Saunji Fyffe, and Jeremy Koulish. 2014. *Data Use for Continuous Quality Improvement: What the Head Start Field Can Learn from Other Disciplines: A Literature Review and Conceptual Framework.* OPRE Report 2014-77. Washington, DC: US Department of Health and Human Services, Office of Planning, Research and Evaluation.

Hoge, Robert D., Gina M. Vincent, and Laura S. Guy. 2012. "Prediction and Risk/Needs Assessments." In *From Juvenile Delinquency to Adult Crime*, edited by Rolf Loeber and David P. Farrington, 150–83. New York: Oxford University Press.

Howell, James C., and Mark W. Lipsey. 2012. "Research-Based Guidelines for Juvenile Justice Programs." *Justice Research and Policy* 14 (1): 17–34.

Lipsey, Mark W. 2009. "The Primary Factors that Characterize Effective Interventions with Juvenile Offenders: A Meta-Analytic Overview." *Victims and Offenders* 4 (2): 124–47.

Lipsey, Mark W., James C. Howell, Marion R. Kelly, Gabrielle Chapman, and Darin Carver. 2010. *Improving the Effectiveness of Juvenile Programs: A New Perspective on Evidence-Based Practice.* Washington, DC: Georgetown University, Center for Juvenile Justice Reform.

Metz, Allison, and Leah Bartley. 2012. "Active Implementation Frameworks for Program Success: How to Use Implementation Science to Improve Outcomes for Children." *Zero to Three* 32 (4): 11–18.

Paulsell, Diane, Kathryn Tout, and Kelly Maxwell. 2013. "Evaluating Implementation of Quality Rating and Improvement Systems." In *Applying Implementation Science in Early Childhood Programs and Systems*, edited by Tamara Halle, Allison Metz, and Ivelisse Martinez-Beck, 269–93. Baltimore: Brookes Publishing.

Redpath, David, and Jeanne K. Brandner. 2010. *The Arizona Standardized Evaluation Protocol (SPEP) for Assessing the Effectiveness of Programs for Juvenile Probationers: SPEP Rating and Relative Recidivism Reduction.* Phoenix: Arizona Supreme Court, Administrative Office of the Courts, Juvenile Justice Service Division. http://cjjr.georgetown.edu/wp-content/uploads/2015/01/pres_AZspep2010.pdf.

Tout, Kathryn, Rebecca Starr, Margaret Soli, Shannon Moodie, Gretchen Kirby, and Kimberly Boller. 2010. *Compendium of Quality Rating Systems and Evaluations.* Washington, DC: US Department of Health and Human Services, Administration for Children and Families, Office of Planning, Research and Evaluation.

Zellman, Gail L., and Michal Perlman. 2008. *Child-Care Quality Rating and Improvement Systems in Five Pioneer States: Implementation Issues and Lessons Learned.* Santa Monica, CA: RAND Corporation.

About the Authors

Akiva Liberman is a senior fellow in the Justice Policy Center at the Urban Institute, where he researches and evaluates crime and justice policy, with a focus on juvenile delinquency and juvenile justice. His current projects include the OJJDP's JJ Reform and Reinvestment Demonstration Program, an evaluation of a juvenile reform demonstration effort centered on evidence-based practices; the Juvenile Second Chance Act Reentry Demonstration Projects, an evaluation of reentry programs for returning juvenile delinquents; and Early Access to Medicaid as a Reentry Strategy, which researches efforts to enroll inmates in Medicaid before release to the community.

Jeanette Hussemann is a research associate in the Justice Policy Center. She is committed to using rigorous social science methods to inform and improve justice systems. Her research portfolio focuses on indigent defense, victimization, and juvenile justice. Hussemann directs projects that employ both quantitative and qualitative research methodologies, and she has extensive experience conducting fieldwork in diverse settings.

Statement of Independence

The Urban Institute strives to meet the highest standards of integrity and quality in its research and analyses and in the evidence-based policy recommendations offered by its researchers and experts. We believe that operating consistent with the values of independence, rigor, and transparency is essential to maintaining those standards. As an organization, the Urban Institute does not take positions on issues, but it does empower and support its experts in sharing their own evidence-based views and policy recommendations that have been shaped by scholarship. Funders do not determine our research findings or the insights and recommendations of our experts. Urban scholars and experts are expected to be objective and follow the evidence wherever it may lead.

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice

2100 M Street NW
Washington, DC 20037

www.urban.org

This resource was prepared by the author(s) using Federal funds provided by the U.S. Department of Justice. Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice