

## A Validation of SPEP™ in Pennsylvania

### A Brief Summary

Edward P. Mulvey  
Carol A. Schubert

### **Background**

The Standardized Program Evaluation Protocol (SPEP™) is a method for determining how well aspects of an existing intervention (or service) match the characteristic of those interventions found to be most effective in reducing recidivism (based on extensive analyses of the research evidence). Under the auspices of the EPISCenter, the implementation of SPEP™ in Pennsylvania has been ongoing since 2012 and the number of different services examined with the SPEP™ has steadily increased over that time. To date, however, there has not been a systematic, large-scale effort to see if these efforts have been “paying off” for key stakeholders in this initiative. Stakeholder groups need some answers:

- the *Commonwealth of Pennsylvania* needs to know if the money provided by PCCD to support this work has improved service quality and outcomes of youth
- *service providers* need to know if the time and effort committed to going through the SPEP™ process makes any difference, and
- they need to be assured that *the youth* served have improved recidivism outcomes as a result of program improvements related to SPEP™

With the support of PCCD, an evaluation team from the University of Pittsburgh (under the direction of Dr. Edward Mulvey) was asked to lead efforts to validate SPEP™ in Pennsylvania with the cooperation of the EPISCenter and JCJC. The EPISCenter provided the evaluation team with data related to elements of services assessed with the SPEP™ protocol (e.g., service type, SPEP™ Total and POP Scores, % of high-risk youth served). JCJC provided de-identified information (e.g., prior court involvement, recidivism outcomes) about youth who were involved in the service at the time of the SPEP™ assessment. Importantly, there was information of this type provided for a subset of services which had both an initial SPEP™ and a reassessment.

This summary document begins with a summary of the primary findings of the validation study but it's important to remember that multiple research questions were examined and a variety of analytic methods were applied. A summary of the procedural details is provided after the conclusions.

### **Summary of Findings**

Arguably the most innovative and important finding from this validation study involves a consideration of how the SPEP™ Total Score changes between the initial assessment and the

first reassessment. *This is the first time any study of the SPEP™ has been able to look at this change.*

➤ **Overall there was improvement in SPEP™ Total Scores for services having an initial and reassessment rating, and positive change was associated with an improved six-month recidivism outcome**

The availability of data on SPEP™ reassessment scores for some services allowed for tests of whether improved or lowered scores were related to recidivism. These analyses indicated some favorable finding about the SPEP™ process as well.

- Thirty-eight services had both an initial and reassessment rating and on average the SPEP™ Total Score improved by 3.16 points. However, 45% of the services had the same or a lower SPEP™ Total Score upon reassessment.
- There was substantial variation in the number of days (range 170 to 1,714) between the date of the initial feedback report review and the SPEP™ reassessment. Unfortunately, the current, limited data is not sufficient to examine the impact of elapsed time between assessments as a factor in the effect of the change scores, but this topic seems worthy of further consideration. Such work could help establish guidelines for the optimal time for reassessments to ensure the consistency and validity of comparisons in program improvement.
- Analyses of the change in scores over time indicated that the amount of change in the SPEP™ Total Score for a service was significantly related to recidivism differences at the six-month point, but not at the twelve-month point. This is an initial finding (since the number of cases is relatively small) indicating that *when a service makes improvements to align with the SPEP™, there is a significant reduction in the six-month recidivism outcome for youth who completed the improved service compared to those who completed the service at its lower level of performance.*

The study also examined patterns and outcomes using just the initial SPEP™ assessment. This work was important to reaffirm the trends and conclusions that were reached when data from other states were examined and it adds to the body of evidence regarding the performance of the SPEP™.

➤ ***As implemented currently statewide, the SPEP™ rating process is producing seemingly valid scores across a variety of services.***

The rating system as applied in Pennsylvania appears to be doing an adequate job of differentiating among services. The patterns of the relationship between the SPEP™ Total and POP Scores and the recidivism outcomes indicated that services with higher SPEP™ scores have better than expected recidivism outcomes for recipient youth and those with low SPEP™ scores have worse than expected recidivism outcomes (variability in SPEP™ scores in the middle range do not show a strong relationship to recidivism).

➤ ***There are discernable subgroups of scores within the full continuum of SPEP™ Total and POP Scores that are associated with recidivism outcomes.***

Analyses were conducted to see if particular subgroups of SPEP™ Total and POP Scores are associated with better recidivism outcomes; the goal of this work was to give service providers benchmarks for achievement of SPEP™ scoring. Three data-driven, score-based groups emerged from this work (SPEP™ Total Score low group=scores 23-43; middle group=scores 44-77; and high group=scores 80-100). These subgroups are associated with better recidivism rates at twelve months. Working toward moving up to the next group thus seems to be a valid goal for service providers.

➤ ***Dimensions of program operations demonstrate varying influence on recidivism outcomes.***

A series of analyses examined the relations between each of the five dimensions of program operations (primary service type, theoretical orientation, evidence-base, setting, and amount/quality of service) and recidivism outcomes. The question examined was whether the difference between observed and expected recidivism is related to the “types” identified in each dimension of service (e.g., locally developed vs. evidence-based practice (EBP); residential vs. community-based) or the score for certain aspects of service provision (i.e., quality, duration, and dosage). The dimensions were tested one at a time to look at their independent effect in this study, not all together (meaning, we did not look at each dimension while simultaneously considering the scores on the other dimensions) as in some of the prior studies.

- The most consistent and strong findings were associated with service setting (community vs. residential). The recidivism differences were significantly more favorable for the community-based services at both six and twelve months. Given that the outcome measure for recidivism incorporates the risk level of the cohort of youths served, this result is not the product of residential services just simply working with “more difficult” adolescents.
- Locally-developed and EBPs were not significantly different from one another in their effect on recidivism outcomes at 6 and 12 months; neither one exerted a strong influence.
- Higher dosage and quality scores have a relationship to the recidivism outcome at twelve months.

The study findings must be viewed in the context of the sample and methods which produced them. In the next few paragraphs we provide some characteristics of the sample and important definitions.

## **The Sample**

This study examined 162 cohorts of youths who received a service that was assessed using SPEP™ between December 2012 and April 2018. Each cohort had 31 youths on average. The services examined represented a variety of service types, with cognitive-behavioral therapy (CBT) programs being the largest group (50 cohorts, 31%) followed by group counseling (17 cohorts, 11%), individual counseling (16 cohorts, 10%), family counseling (14 cohorts, 9%), job-related training and social skills training (each with 11 cohorts, 14%). There were 2,496 distinct youth represented in the cohorts (some youth were in more than one cohort). Thirty-eight of these cohorts also had a SPEP™ reassessment. The reassessment sample is composed of services from primarily three service types: cognitive behavioral programs (n = 14; 37%), family counseling (n = 6; 16%), and behavioral contracting (n = 4; 11%). A little over a thousand youths (N = 1,037) were in the cohorts connected to these services.

## **The Outcome**

Recidivism is the primary outcome of interest. For the purposes of this validation study, we define a recidivating event as either an adjudication or conviction for a misdemeanor or felony offense within two time periods: 1) six months (1-180 days) after the SPEP™ service end date or 2) twelve months (1-365 days) after the service end date. Using court record information provided by JCJC, each youth in the sample was given a positive value for a “*six month observed recidivism*” or “*twelve month observed recidivism*” if they met the above definition. A “*cohort observed recidivism rate*” was calculated for six and twelve months as well; this is simply the percent of youth in the cohort meeting each definition.

The youths in the sample are not all the same. They have a range of background characteristics that put them at more or less risk to recidivate. Also, these youths end up in particular services for a variety of reasons (e.g., court mandate, available beds, service goals). This means that some services are likely to have more high-risk youth than other services. Given this, simply comparing observed recidivism rates across services is not a fair assessment of effectiveness.

To account for the variability in risk (and because YLS scores were not available for a large proportion of the sample), a score was calculated for each youth to reflect his/her likelihood of recidivating during the six month and/or 12 months after their involvement with the service (we call this the *expected recidivism risk* for that individual). This score is based on the characteristics of the adolescent at the time of entry into the service being assessed; the characteristics used to derive this score (e.g., age at first offense, number of prior contacts with the juvenile justice system) are evidence-based, well-established risk factors for recidivism. These expected recidivism scores showed the same level of predictive accuracy as other risk assessment tools and corresponded well to the scores derived from the YLS/CMI. A “*cohort expected recidivism rate*” was then calculated for six and 12 months as the average predicted recidivism risk of youth in the cohort.

The *cohort observed recidivism rate* and the *cohort expected recidivism rate* are compared to determine if a particular service cohort is performing better or worse than might be expected, given the risk level of the group of youth served. This is done by simply subtracting the expected rate of the cohort from the observed rate of the cohort. If the expected rate is higher than the observed rate, the service is performing better than would have been expected given the risk of the youth served (the group recidivism rate is lower than what would have been predicted). If the expected rate is lower than the observed rate, the service is performing worse (the observed rate is higher than would have been expected).

Overall, it appears that efforts to implement the SPEP™ in Pennsylvania are well placed and generally successful. This study provides preliminary, but clear and consistent, support for the idea that program improvements in the SPEP™ framework should reduce recidivism by systematically improving service provision.

More detail about the study methods and finding can be found at:  
<https://www.pccd.pa.gov/Juvenile-Justice/Pages/Standardized-Program-Evaluation-Protocol.aspx>