

**The Arizona Standardized Program Evaluation Protocol (SPEP) for Assessing
the Effectiveness of Programs for Juvenile Probationers:**

SPEP Rating and Relative Recidivism Reduction

An Update to the January 2008 Report by Dr. Mark Lipsey

Prepared by:

**David P. Redpath & Jeanne K. Brandner
Arizona Supreme Court
Administrative Office of the Courts
Juvenile Justice Service Division
Phoenix, AZ**

April 2010

Table of Contents

| | |
|--|----|
| Report..... | 1 |
| Background..... | 1 |
| What is the SPEP? | 1 |
| What Characteristics are rated by the SPEP? | 1 |
| Figure 1: General Form of the SPEP Instrument..... | 2 |
| Purpose of the SPEP | 3 |
| Purpose of this Report..... | 3 |
| Data Available for the Analysis..... | 3 |
| SPEP Scores for Service Categories..... | 3 |
| Table 1: Service Categories for Programs that Served juveniles with at least 6- Month Recidivism Data..... | 4 |
| Figure 2: Number of Programs with Different Total SPEP Scores | 5 |
| Table 2: Number and Percentage of Juveniles Served by Programs with Different SPEP Scores (Including Program Numbers and Percentages)..... | 5 |
| Table 3: SPEP Component Scores..... | 6 |
| Recidivism of the Juveniles Served | 6 |
| Twelve Month Recidivism..... | 6 |
| Six Month Recidivism | 7 |
| The Difference between Predicted and Actual Recidivism | 7 |
| Recidivism for Service Providers with SPEP Scores | 8 |
| Figure 3: Difference between Actual and Predicted Recidivism for Service Providers with SPEP Scores of 45 or More vs. Scores of Less than 45..... | 9 |
| Table 4: Correlations across Service Providers of the SPEP Component and Total Scores with the Difference between Actual and Predicted Recidivism | 10 |
| Conclusions..... | 11 |
| Appendix A: Computing Predicted Recidivism | 13 |
| Table A1: Logistic Regression Results for Predicting 12-Month Recidivism | 14 |

Report

Background

The Standardized Program Evaluation Protocol (SPEP) is a rating instrument for assessing programs for juvenile offenders with regard to their expected effectiveness for reducing recidivism. The Administrative Office of the Courts (AOC) Juvenile Justice Services Division (JJSD) and the Peabody Research Institute at Vanderbilt University partnered to adapt a version of the SPEP for evaluation of Arizona programs serving juvenile probationers. A team of national experts, under the direction of Dr. Mark Lipsey, Director of the Peabody Research Institute at Vanderbilt, assisted with the SPEP implementation process. In the fall of 2006, Arizona began implementing the SPEP rating scheme for contracted service providers in five (Maricopa, Yavapai, Yuma, Pima, and Pinal) pilot counties. In 2007, Lipsey conducted a preliminary investigation of whether the SPEP ratings of service programs were related to recidivism outcomes for the juveniles they served. He found that juveniles served by providers with higher SPEP scores (>50 points) had recidivism rates 12-13 percent lower than predicted on the basis of their assessed level of risk, while juveniles served by providers with lower SPEP scores (<49 points) recidivated at rates within one percentage point of what was predicted. In the summer of 2007, Arizona expanded the SPEP to all Arizona counties. This report serves as an update to Lipsey's preliminary investigation, with the inclusion of details from the ten remaining counties. Recent data renders updated SPEP scores for providers in the pilot counties and initial SPEP scores for providers in the ten additional counties. Lastly, recidivism data for many more juveniles served by providers who have been evaluated by the SPEP are included for analysis.

What is the SPEP?

The SPEP is based on research on the effectiveness of programs for juvenile offenders drawn from an archive of nearly 600 controlled studies of effects on recidivism assembled by Lipsey. Using the technique of meta-analysis, a sophisticated way to extract, analyze, and summarize findings of a group of related research studies, the characteristics of programs with the largest impacts on recidivism have been identified from research and translated into guidelines for effective interventions. The SPEP uses these guidelines as a framework for developing ratings of programs according to how closely their characteristics match those shown in research to be most strongly associated with recidivism reduction.

What Characteristics are rated by the SPEP?

The program characteristics rated by the SPEP include: the primary service category, supplemental services, duration and frequency of service, quality of service, and the delinquency risk level of juveniles served. Figure 1 is an example of a SPEP score card, representing the rating categories and scoring scheme of the SPEP instrument. Individual programs provided by contracted vendors receive a SPEP score generated from records extracted by JJSD that provide data on the type and amount of service and the risk assessment scores for the juveniles served. The component of program quality is not currently included in these scores although JJSD is looking to add this feature in the next round of scores released. Currently, SPEP scores can range from 15 to 85 points.

Figure 1: General Form of the SPEP Instrument

| AZ SPEP for Services to Probation Youth | | |
|---|------------------------|------------------------|
| Primary Service: | Possible Points | Received Points |
| High Effect (35 points) Moderate Effect (25 points) Low Effect (15 points) | 35 | |
| Supplemental Service: | 5 | |
| Qualifying supplemental service used (5 points) | | |
| Treatment Amount: | 10 | |
| Duration: % of youth that received target weeks of service or more 0% (0 points) 60% (6 points) 20% (2 points) 80% (8 points) 40% (4 points) 100% (10 points) | | |
| Contact Hours: % of youth that received target hours of service or more 0% (0 points) 60% (9 points) 20% (3 points) 80% (12 points) 40% (6 points) 100% (15 points) | | |
| Treatment Quality: <i>THIS IS NOT YET EVALUATED</i> Low Medium High | [15] | NOT SCORED |
| Youth Delinquency Risk Level: | 20 | |
| % of youth with a risk score over .50 (medium) < 85% (0 points) 85% (5 points) 99% (10 points) High Risk: % of youth with a risk score over .70 (high) < 70% (0 points) 70% (5 points) 90% (10 points) | | |
| Provider's Total SPEP Score: | 100 | [TOTAL SCORE] |

Purpose of the SPEP

The SPEP rating scheme assigns points for the respective program characteristics according to research findings about their relationship to recidivism. Therefore, the overall SPEP score of each program represents an evaluation of its expected effectiveness for reducing recidivism. Lower than maximum ratings for each different program characteristic signifies an area in which a program has potential for improvement. The purpose of the Arizona SPEP, therefore, is to (a) evaluate the effectiveness of individual programs provided by JJSD contracted vendors for juvenile probation youth and (b) guide vendors towards improvements that will make those programs more effective.

Purpose of this Report

As previously mentioned, this report serves as an update to Lipsey's preliminary investigation. While Lipsey summarized data from five pilot counties (Maricopa, Yavapai, Yuma, Pima, and Pinal), this report expands the data to all fifteen Arizona counties analyzing the SPEP scores for providers funded by JJSD for services to juvenile probation youth, the recidivism of those juveniles after receiving services, and the relationship of SPEP scores to recidivism. The first two of these components are mainly descriptive, looking at SPEP scores for providers and recidivism for juveniles. The third component, however, addresses the validity of the SPEP scores. In other words, if SPEP scores are indeed useful guides to effective programming, there should be a correlation for juveniles of similar risk level between their recidivism rates and the SPEP scores of the programs from which they received services.

Data Available for the Analysis

The Juvenile Online Tracking System (JOLTS) is Arizona's statewide juvenile probation and dependency management system developed by Maricopa County Juvenile Court in 1979. The JOLTS is currently utilized in every juvenile court and detention center in Arizona, with the exception of Maricopa County which utilizes a parallel database, Integrated Court Information System (iCIS). Data used for this report included (a) therapeutic service provision/billing records for juvenile probationers who completed service between February 2005 and November 2007, and contributed to the SPEP scores for the providers of those services, combined with (b) JOLTS/iCIS offense records for those juveniles through August 2008. For some juveniles there were multiple therapeutic service records. Receiving multiple services creates uncertainty surrounding the extent to which recidivism was influenced by the most recent service. Records were therefore dropped from the analysis for juveniles with more than three service episodes; for the remainder, the most recent service episode was used. Records were also dropped for juveniles who did not have sufficient time past the end of service for either the 6-month or 12-month recidivism data to accumulate and those whose age at the conclusion of service did not leave the respective 6- or 12-month interval before they turned 18.

SPEP Scores for Service Categories

Juveniles with 6-month recidivism data available for analysis were served by 90 SPEP rated programs. Table 1 exhibits the service categories into which the programs were classified by JJSD staff, the number of programs in each category, and the number of juveniles with recidivism data served by each program type.

Table 1: Service Categories for Programs that Served juveniles with at least 6-Month Recidivism Data

| Service Category | Number of Programs | Number of Juveniles |
|-----------------------------------|---------------------------|----------------------------|
| Cognitive Behavioral, Community | 1 | 62 |
| Cognitive Behavioral, Residential | 3 | 162 |
| Family Counseling, FFT | 1 | 106 |
| Family Counseling Therapy | 11 | 811 |
| Group Counseling, Community | 9 | 394 |
| Group Counseling, Residential | 5 | 127 |
| Individual Counseling | 18 | 575 |
| Life Skills Training | 2 | 89 |
| Mentoring | 2 | 218 |
| Sex Offender, Community | 8 | 128 |
| Sex Offender, Residential | 4 | 52 |
| Substance Abuse, Community | 25 | 832 |
| Substance Abuse, Residential | 1 | 15 |
| TOTALS | 90 | 3571 |

For SPEP rated programs, higher SPEP scores identify programs expected to be more effective on the basis of available research on similar programs. The total SPEP scores, as currently computed, are composed of three main component scores, each further divided into two subparts, as follows:

- Type of service (primary service and qualifying supplementary services);
- Amount of service (service duration in weeks and total number of contact hours);
- Delinquency risk level of juveniles served (proportions reaching moderate and high).

In addition, there is a fourth component for quality of service, worth up to 15 points, which is not currently rated. The maximum possible SPEP score for programs in the present analysis, therefore, is 85.

Figure 2 reports the distribution of total SPEP scores for the programs that served juveniles for whom 6-month recidivism data were available. The most notable aspect of the data in Figure 2 is that there are relatively few programs scoring on the high end. In Lipsey’s original analysis of the five pilot counties, there was also a relatively low and restricted range of total SPEP scores for programs. Lipsey found that, of a maximum possible total score of 85, 73% of the providers scored under 50. The statewide data shows very similar results, 58.8% of the programs scoring under 45 and 86.6% scoring under 54 (these percentages are illustrated in Table 2). In other words, for most of the programs there continues to be ample room for improvement.

Figure 2: Number of Programs with Different Total SPEP Scores

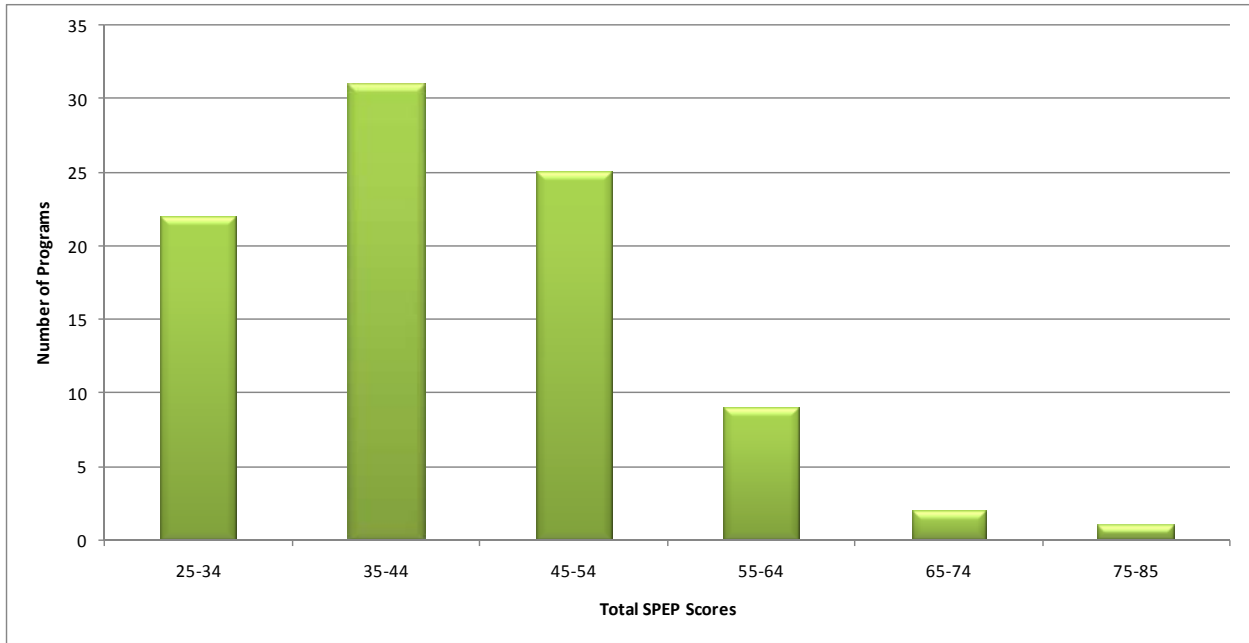


Table 2 shows the number of juveniles served by providers in the different SPEP score ranges. This table further emphasizes that programs have ample room for improvements. About 56% of juveniles in a SPEP rated program were in programs with ratings of less than 45. On the other end of the spectrum, approximately 5% of juveniles were in programs with ratings of 65 or higher.

Table 2: Number and Percentage of Juveniles Served by Programs with Different SPEP Scores (Including Program Numbers and Percentages)

| SPEP Score | N Juveniles | Juveniles % | N Programs | Programs % |
|-------------------|--------------------|--------------------|-------------------|-------------------|
| 25-34 | 1049 | 29.4% | 22 | 24.4% |
| 35-44 | 959 | 26.9% | 31 | 34.4% |
| 45-54 | 977 | 27.4% | 25 | 27.8% |
| 55-64 | 403 | 11.3% | 9 | 10.0% |
| 65-74 | 63 | 1.8% | 2 | 2.2% |
| 75-85 | 120 | 3.4% | 1 | 1.1% |
| TOTALS | 3571 | 100.0% | 90 | 100.0% |

Table 3 reports the distributions of component scores that went into the total SPEP scores for the SPEP rated programs. These breakdowns reveal that the greatest shortfall was for amount of service, where over a third of the providers scored zero and more than 90% received fewer than half the 25 points available in this category. The next largest shortfall was for juvenile risk with 79% of the providers receiving half or fewer of the 20 points available. The strongest category was type of service with only 11% of the providers scoring half or fewer of the 40 points available.

Table 3: SPEP Component Scores

| Service Type (Max 40) | N | % | Primary (Max 35) | N | % | Supplemental (Max 5) | N | % | |
|--------------------------|-----------|---------------|---------------------|----|-----------|-------------------------|---|-----------|---------------|
| 15 | 1 | 1.1% | | 15 | 11 | 12.2% | 0 | 62 | 68.9% |
| 20 | 9 | 10.0% | | 25 | 38 | 42.2% | 5 | 28 | 31.1% |
| 25 | 29 | 32.2% | | 35 | 41 | 45.6% | | | |
| 30 | 9 | 10.0% | | | | | | | |
| 35 | 31 | 34.4% | | | | | | | |
| 40 | 11 | 12.2% | | | | | | | |
| | <u>90</u> | <u>100.0%</u> | | | <u>90</u> | <u>100.0%</u> | | <u>90</u> | <u>100.0%</u> |

| Service Amount (Max 25) | N | % | Weeks (Max 10) | N | % | Hours (Max 15) | N | % | |
|----------------------------|-----------|---------------|-------------------|----|-----------|-------------------|----|-----------|---------------|
| 0 | 34 | 37.8% | | 0 | 36 | 40.0% | 0 | 60 | 66.7% |
| 2-4 | 25 | 27.8% | | 2 | 18 | 20.0% | 3 | 7 | 7.8% |
| 5-7 | 8 | 8.9% | | 4 | 26 | 28.9% | 6 | 9 | 10.0% |
| 8-10 | 7 | 7.8% | | 6 | 9 | 10.0% | 9 | 8 | 8.9% |
| 11-13 | 8 | 8.9% | | 8 | 1 | 1.1% | 12 | 6 | 6.7% |
| 14-16 | 5 | 5.6% | | 10 | 0 | 0.0% | 15 | 0 | 0.0% |
| 17-20 | 3 | 3.3% | | | | | | | |
| 20-25 | 0 | 0.0% | | | | | | | |
| | <u>90</u> | <u>100.0%</u> | | | <u>90</u> | <u>100.0%</u> | | <u>90</u> | <u>100.0%</u> |

| Client Risk (Max 20) | N | % | Moderate (Max 10) | N | % | High Level (Max 10) | N | % | |
|-------------------------|-----------|---------------|----------------------|----|-----------|------------------------|----|-----------|---------------|
| 0 | 32 | 35.6% | | 0 | 36 | 40.0% | 0 | 39 | 43.3% |
| 5 | 10 | 11.1% | | 5 | 42 | 46.7% | 5 | 33 | 36.7% |
| 10 | 29 | 32.2% | | 10 | 12 | 13.3% | 10 | 18 | 20.0% |
| 15 | 9 | 10.0% | | | | | | | |
| 20 | 10 | 11.1% | | | | | | | |
| | <u>90</u> | <u>100.0%</u> | | | <u>90</u> | <u>100.0%</u> | | <u>90</u> | <u>100.0%</u> |

Recidivism of the Juveniles Served

The following recidivism variables were used in the analysis:

- 12-month recidivism—whether any new complaint was recorded for a delinquency or status offense in the twelve months after the end of service (yes/no);
- 6-month recidivism—whether any new complaint was recorded for a delinquency or status offense in the six months after the end of service (yes/no).

It should be noted that the following recidivism rates are for the select group of juveniles who appear in the analysis sample and are not representative of the recidivism rates for the whole population of juveniles served by the Arizona juvenile justice system. The juveniles in the analysis sample are only those who received services paid by JJSD during the restricted time windows that allowed for 6- and 12-month recidivism data to accumulate.

Twelve Month Recidivism

The data in the 12-month recidivism analysis sample for the fifteen counties statewide included 2681 juveniles who were at least 12 months short of their 18th birthday and whose offense records extended at least 12 months past the date on which their service ended. The average age of these juveniles at the end of service was 15.6; 75% were male, 48% Caucasian, 40% Hispanic, 7% African-American, 4% Native American, and 1% Asian. The 6-month recidivism rate for these juveniles was .23 and the 12-month recidivism rate was .38. That is, 23% and 38% of the juveniles, respectively, had a new complaint for a delinquency or status offense following the six or twelve months after the end of service.

Six Month Recidivism

An additional 890 juveniles were at least 6 months before their 18th birthday and had offense records for 6 months following the end of service (but not 12 months). Combined with the 2681 juveniles above, this made a total sample of 3571 juveniles for whom 6-month recidivism data were available. Their 6-month recidivism rate was the same as above, .23.

The Difference between Predicted and Actual Recidivism

Recidivism rates are largely a function of risk factors present before receipt of services, so it can be very misleading to compare programs with regard to the recidivism rates of the juveniles they serve. A program may be relatively ineffective but serve low risk juveniles who, therefore, have the same low recidivism rate they would have had without the program intervention. Another program may serve high risk juveniles, who would otherwise have a high recidivism rate, and may be very effective in reducing their recidivism. Despite the effectiveness of the program, however, the juveniles could still have a recidivism rate higher than a program that started with low risk juveniles. A relatively high proportion of high risk juveniles will recidivate despite effective services, just not as many as with ineffective services.

Under these circumstances, the only way to assess the effects of program services on recidivism with any confidence is through controlled research in which juveniles are randomly assigned to different services and their subsequent recidivism compared. Absent results from such research, the best alternative is to attempt to statistically adjust recidivism rates for prior risk using data on the juveniles' characteristics before they received service. Such adjustments amount to attempting to predict recidivism on the basis of the juveniles' prior risk-related characteristics. If those adjustments are successful, the difference between predicted recidivism and actual recidivism will then be an indication of the effectiveness of service. More effective programs should show an actual recidivism rate lower than the predicted recidivism.

The first step in the analysis of the recidivism rates for the various programs, therefore, was to identify variables relating to juveniles' prior status that were capable of statistically predicting recidivism. These variables were then used to generate a predicted probability of recidivism for each juvenile so that, for juveniles receiving different services, their actual recidivism could be compared with their predicted recidivism. The variables selected for this analysis, based on their significant contributions to the prediction of recidivism, included prior service events, county, age, sex, race, risk rating, and the number and nature of prior complaints recorded. Further details about the variables and the statistical prediction procedure are presented in Appendix A.

The predicted probability of recidivism values that resulted from this analysis for 6-month and 12-month recidivism had correlations ranging from .23 to .31 with actual recidivism. These correlations represent a predictive accuracy of about 68%; that is, the predicted recidivism value was correct for about 68% of the cases. These are modest correlations, however, though in the range typically found in validation studies of recidivism risk rating instruments. They are not such high correlations that we can be certain that all the relevant predictive factors are represented. To the extent that important variables are left out of the prediction equation, the resulting predictions may be too high or too low for some juveniles. This ambiguity should be kept in mind when interpreting the comparisons between actual and predicted recidivism for

different service programs. If a program serves juveniles with unmeasured characteristics that make them lower risk than the recidivism prediction indicates, the actual recidivism will be lower because of that prediction error and not necessarily because of the effectiveness of the program. Conversely, unmeasured characteristics that make risk higher than the predicted recidivism indicates will make an effective program look less effective.

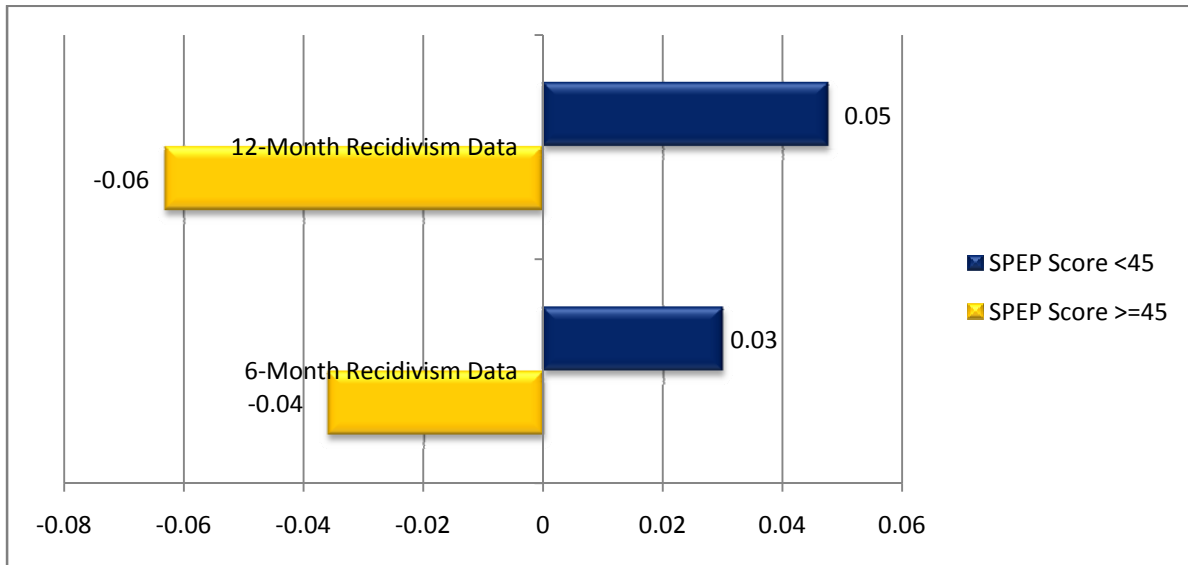
Another factor that should be kept in mind is that virtually all the juveniles in the analysis received at least some service. Thus the predicted recidivism estimate is the expected recidivism *with the average service*—the analysis does not include a “no program” control group that allows estimation of recidivism absent service. The value of these predicted rates is that we expect juveniles receiving a service that is more effective than average to show an actual recidivism lower than predicted. On the other hand, juveniles receiving a service less effective than average should show actual recidivism higher than predicted.

Recidivism for Service Providers with SPEP Scores

SPEP scores rate individual service providers based on the services they provide to their juvenile clients. The most informative analysis of the extent to which SPEP scores are valid indicators of program effectiveness, therefore, is one that examines the relationship between the scores for individual providers and the recidivism of the juveniles they serve. As described earlier, there are limitations to any such analysis that must be kept in mind. The SPEP ratings with which we are working are incomplete—they do not include the important component that represents the quality of the services provided. In addition, our ability to predict the expected recidivism from the available data is modest. Omitted variables in the statistical prediction model can cause over or under prediction with corresponding misestimation of program effectiveness. Additionally, there are a few SPEP categories with rather small numbers of juveniles served, making analysis findings relatively unstable. If the SPEP ratings are valid indications of program effectiveness, however, we would expect the available data to show some indication that service providers with higher SPEP scores have better risk-adjusted recidivism outcomes despite these limitations.

As shown in Figure 2, presented earlier, more than 58% of the providers in the analysis sample had a total SPEP score below 45 (85 is currently the maximum possible score). Figure 3, below, shows the mean difference between actual and predicted recidivism rates for the service providers with SPEP scores of 45 or more compared with the difference for providers with SPEP scores of less than 45 (weighted by the number of cases available for each provider). For both 6- and 12-month recidivism, juveniles served by providers with SPEP scores of 45 or higher had recidivism rates 3 to 7 percentage points lower than predicted on the basis of their prior risk. In contrast, juveniles served by providers with SPEP scores lower than 45 had recidivism rates closer to the predicted values—about 3 to 5 percentage points higher than predicted. For individual providers, therefore, higher SPEP scores were strongly associated with larger reductions in recidivism.

Figure 3: Difference between Actual and Predicted Recidivism for Service Providers with SPEP Scores of 45 or More vs. Scores of Less than 45



The relative magnitude of the SPEP-related differences between actual and predicted recidivism can be illustrated with the 12-month recidivism rates. Juveniles served by providers with SPEP scores of 45 or more had a predicted recidivism rate of .49 and an actual rate of .43, implying a level of program effectiveness that produced a reduction of 6 percentage points in the recidivism relative to the average effects for all services. A .06 point reduction from the expected .49 rate is a 12% decrease in the recidivism rate. The juveniles served by providers with SPEP scores under 45, on the other hand, had a predicted 12-month recidivism rate of .30 and an actual rate of .35 implying a below average level of program effectiveness with an actual recidivism rate 5 percentage points higher than predicted based on the risk level of the juveniles they treated. Put another way, there is an 11 percentage point difference between the high and low scoring providers in how the actual recidivism rates of the juveniles served compare with the rate predicted from the juveniles' risk level. The higher rated providers are more effective than average; while the lower rated providers are less effective than average. While limitations of the data and analysis must be acknowledged, these results give quite positive indications of the validity of the SPEP ratings for identifying programs that are effective in reducing recidivism.

To provide a more detailed analysis, correlations between providers' SPEP scores and their 6-month and 12-month relative recidivism reductions were also computed. The correlations with the total SPEP scores were examined along with those for the SPEP component scores relating to type of service, amount of service, and risk level of the juveniles served (see Table 3, presented earlier, for the distribution of the total and component scores). Within the limits of the data available, these correlations can be viewed as validity coefficients for the SPEP ratings. If the SPEP scores are related to program effectiveness as expected, we should find higher scores associated with recidivism rates lower than predicted from prior risk factors. That is, there should be significant negative correlations between the various SPEP scores and the actual-minus-predicted recidivism differences across service providers. Since the SPEP scores were not

designed as interval scales, nonparametric Spearman rank order correlation coefficients were used for this analysis.

Table 4 shows the resulting correlations. The first thing to notice about these correlations is that nearly all are negative, showing the expected direction of effects between higher SPEP scores and lower than expected recidivism, and some of these are statistically significant. The correlations with the total SPEP score range from -.143 to -.360, with the 12-month recidivism reductions attaining statistical significance. Once again we see that juveniles served by providers with higher overall SPEP scores show lower than predicted recidivism. These results provide good support for the assumption on which the SPEP is based—that Arizona programs with characteristics more closely matching those shown effective in research studies will in fact be more effective in reducing recidivism.

The correlations in Table 5 that involve the component scores that contribute to the total SPEP ratings generally show that each of them is separately associated with relative recidivism reductions. The relationships of the scores for risk are especially consistent—all are in the right direction and all are statistically significant. The component scores for twelve month recidivism are all in the right direction, but not all attain statistical significance.

Table 4: Correlations across Service Providers of the SPEP Component and Total Scores with the Difference between Actual and Predicted Recidivism

| SPEP Scores | 6 Month Recidivism N=67 | 12 Month Recidivism N=57 |
|-----------------------------------|----------------------------|-----------------------------|
| Type of Service Subtotal | .020 | -.178 |
| Primary | .067 | -.136 |
| Supplemental | -.135 | -.218 |
| Amount of Contact Subtotal | -.024 | -.186 |
| Weeks | -.100 | -.122 |
| Hours | .019 | -.209 |
| Risk Subtotal | -.348** | -.470** |
| Lower Risk Range | -.309* | -.454** |
| Upper Risk Range | -.348** | -.422** |
| Total SPEP Score | -0.143 | -.360** |

*p<.05 **p<.01

Sperman's rho correlation

Conclusions

The primary purpose of the analyses summarized in this report was to investigate the validity of the SPEP rating scheme as an evidence-based assessment of the effectiveness of programs for reducing the recidivism of juvenile offenders. Data available were limited to baseline SPEP ratings of providers, some of which were based on relatively modest numbers of juvenile service records for closed cases. Moreover, complete recidivism data were not available for all of the juveniles represented in those service records. SPEP ratings are generated for individual programs of contracted service providers from their service provision/billing records and neither those ratings nor the recidivism rates for the juveniles served by those providers are fully representative of program performance when based on relatively small numbers of closed cases. Finally, as previously mentioned, total SPEP scores are currently lacking the critical component of treatment quality.

Aside from the data limitations, there are inherent limitations in the analyses conducted for this report. Juveniles are not randomly assigned to service providers so we cannot assume that the youth they serve have equal risk of recidivism. Lower recidivism rates, therefore, do not necessarily indicate better program performance—they may only indicate that a provider served lower risk juveniles. To help level the playing field for recidivism comparisons across programs with different SPEP scores, the expected recidivism for the youth served was predicted from a battery of prior risk and demographic factors and actual recidivism was then compared with predicted recidivism. The recidivism rates for programs more effective than average, as evidenced by a higher SPEP score, should be notably lower than those predicted for the juveniles they serve, and the recidivism for programs less effective than average should be the same or higher than predicted. Though this technique should make recidivism outcomes more comparable across programs, it is far from perfect. Recidivism is difficult to predict and we have no assurance that all the relevant risk factors have been accounted for in the procedures used in the analyses reported here. On the other hand, there is little reason to believe that whatever bias remains in the estimated recidivism outcomes is likely to be systematically related to SPEP scores. That would require juveniles less likely to recidivate than predicted to be more likely to be assigned to service providers with high SPEP scores. SPEP scores were not available at the time these juveniles were assigned to services and there are no obvious characteristics of higher scoring services that would stimulate referrals of juveniles with unmeasured favorable risk profiles across the multiple counties and many providers represented in the available data.

In light of these various considerations and limitations, the results of the analyses reported here continue to be encouraging. The SPEP scores do show statistically significant and relatively strong relationships with the risk-adjusted recidivism outcomes for the juveniles served by the respective service providers. ***Juvenile offenders served by providers with higher SPEP scores had lower than predicted recidivism; juveniles served by providers with lower SPEP score recidivated at a rate higher than predicted. The main conclusion of this report therefore, is that the SPEP scores are working as expected and do show promising empirical validity as guides to effective programming for juvenile offenders.***

Given indications that the SPEP ratings have sufficient validity to guide programming, it is relevant to consider how well the Arizona programs score on the SPEP. With a maximum possible score of 85 at present (15 points remain for service quality, but this component is not yet

scored in the SPEP), more than 58% of the programs that could be rated scored below 45, with most between 35 and 44. It is not surprising that real world programs do not match the characteristics of the best programs represented in research studies of program effects, so these ratings should not be taken to indicate that the Arizona programs are especially weak. Nonetheless, there is clearly room for improvement and the preliminary validity results for the SPEP ratings suggest that improving the SPEP scores should translate into greater effects on recidivism. The breakdown of the SPEP component scores shows that the Arizona programs are generally using effective types of services and serving some juveniles of sufficiently high risk to warrant service. Though there is room for improvement in these areas, especially with regard to the targeting of high risk juveniles, the greatest shortfall in the SPEP scoring appears for the amount of service—the duration of service and, especially, the number of service contact hours with the juveniles served. These components of the SPEP showed promising correlations with relative recidivism reductions so there is every reason to believe that increases in the amount of service provided will yield larger effects on recidivism. *The second major conclusion of this report, therefore, is that there is ample room for improvement in the effectiveness of the Arizona programs as indexed by the SPEP ratings. Increases in the amount of service provided along with more focus on high risk cases, while maintaining an emphasis on the more effective types of service, is likely to yield the largest effects on recidivism reduction.*

Arizona's JJSD has vigorously implemented the SPEP assessment scheme and has firmly and constructively encouraged service providers to use those assessments to guide program improvements. Efforts in this regard are in large part motivated by the belief that this process will lead to better use of the programs currently available, incremental development of more effective programs, and, ultimately, lower recidivism rates and greater public safety. Lipsey's preliminary report and JJSD's follow-up reveal encouraging findings in support of the potential value of the SPEP assessment for identifying effective programs and providing guidance for program improvements.

Appendix A: Computing Predicted Recidivism

Predicted recidivism values for each juvenile were created with a logistical regression analysis in which a selected set of predictor variables was regressed, in separate analyses, on the dichotomous 6-month and 12-month recidivism variables. The following variables were found collectively to be the best predictors:

- Number of prior service events (about 25% of the juveniles had one or two service events prior to the one analyzed; those with more were dropped from the analysis);
- Total cumulative months of prior service;
- County—dummy codes for all counties with county 8 omitted as the reference group;
- Age at end of service;
- Sex;
- Race—dummy codes for Black, Hispanic, Native Americans, and Asian with White omitted as the reference group;
- Risk rating from the Risk Assessment Instrument completed by the probation officer (with maximum likelihood imputation of the missing values);
- Selected individual items from the Risk Rating instrument-- Juvenile's Relationship with Family, Drug Involvement, Truancy, School, Runaway, Probation Officers' ratings of the likelihood of re-offense;
- Age at first offense;
- Log of the number of total prior complaints (including VOP);
- Log of the prior VOP complaints;
- Log of the number of prior VOP counts;
- Mean Severity index of prior offenses (Type x Class codes rescaled);
- Mean Severity of the prior complaint dispositions;
- Total number of prior days detained.

These predictors were used in separate logistic regression analyses to predict 6-month and 12-month recidivism. The predicted values (predicted probability of recidivism) that resulted from these analyses were then dichotomized at the points that provided the same recidivism proportions as the actual recidivism being predicted and recoded for each juvenile as 1 if recidivism was predicted and 0 if it was not. That dichotomized variable was then used to generate the predicted recidivism rates for the various groups of juveniles examined in the analysis. The correlations of the predicted and actual recidivism values in each case were statistically significant with correlation coefficients as follows:

- 6-month recidivism: .23 (correctly predicted for 71% of the cases);
- 12-month recidivism: .31 (correctly predicted for 65% of the cases).

As an example, the results of the logistic regression analysis for predictors of 12-month recidivism are presented in Table A1 below. This analysis includes 5167 records. The overall prediction model is statistically significant with a Chi-Square of 492.2 (df=36), $p < .001$. It is notable that one of the best single predictors in this model is the risk score from the Risk Assessment instrument completed by the probation officers and used in the SPEP ratings. (The PO's risk ratings by themselves correlate .16 with 6-month recidivism and .18 with 12-month recidivism).

Table A1: Logistic Regression Results for Predicting 12-Month Recidivism

| | B | S.E. | Wald | df | Sig. | Exp(B) |
|---------------------|--------|------|--------|----|------|--------|
| Step 1 ^a | | | | | | |
| priorservice_number | .085 | .087 | .946 | 1 | .331 | 1.089 |
| priorservice_months | -.029 | .023 | 1.605 | 1 | .205 | .971 |
| County01 | -.046 | .453 | .010 | 1 | .919 | .955 |
| County02 | .349 | .200 | 3.063 | 1 | .080 | 1.418 |
| County03 | .202 | .202 | .998 | 1 | .318 | 1.224 |
| County04 | .331 | .230 | 2.071 | 1 | .150 | 1.392 |
| County05 | .552 | .222 | 6.164 | 1 | .013 | 1.736 |
| County06 | -.105 | .471 | .050 | 1 | .824 | .900 |
| County07 | -.197 | .517 | .145 | 1 | .703 | .821 |
| County09 | .177 | .151 | 1.368 | 1 | .242 | 1.193 |
| County10 | -.292 | .358 | .666 | 1 | .414 | .747 |
| County11 | .481 | .118 | 16.729 | 1 | .000 | 1.618 |
| County12 | -.369 | .143 | 6.624 | 1 | .010 | .691 |
| County13 | .071 | .252 | .080 | 1 | .778 | 1.074 |
| County14 | -.791 | .172 | 21.066 | 1 | .000 | .454 |
| County15 | .098 | .134 | .531 | 1 | .466 | 1.102 |
| age | -.081 | .034 | 5.670 | 1 | .017 | .922 |
| R_Sex | .696 | .078 | 79.133 | 1 | .000 | 2.006 |
| Race_black | .313 | .113 | 7.625 | 1 | .006 | 1.367 |
| Race_hispanic | .227 | .070 | 10.631 | 1 | .001 | 1.254 |
| Race_indian | .097 | .170 | .326 | 1 | .568 | 1.102 |
| Race_asian | -.010 | .407 | .001 | 1 | .981 | .990 |
| RiskScore | 1.377 | .326 | 17.811 | 1 | .000 | 3.963 |
| Juvrel | .000 | .075 | .000 | 1 | .995 | 1.000 |
| Drugs | .116 | .082 | 2.034 | 1 | .154 | 1.123 |
| Truancy | -.143 | .078 | 3.358 | 1 | .067 | .867 |
| School | -.095 | .066 | 2.042 | 1 | .153 | .909 |
| Runaway | -.065 | .081 | .641 | 1 | .423 | .937 |
| Reoffend | .137 | .055 | 6.228 | 1 | .013 | 1.147 |
| P_CAge_First | .066 | .022 | 8.863 | 1 | .003 | 1.068 |
| log_NTotComplaints | .620 | .116 | 28.364 | 1 | .000 | 1.860 |
| log_VOPComplaints | -.169 | .143 | 1.402 | 1 | .236 | .845 |
| log_VOPcts | -.101 | .083 | 1.460 | 1 | .227 | .904 |
| P_CSev_index_mean | -.006 | .014 | .165 | 1 | .685 | .994 |
| P_CDispo_Sev_mean | .000 | .037 | .000 | 1 | .993 | 1.000 |
| P_Days_Detained | -.002 | .001 | 9.132 | 1 | .003 | .998 |
| Constant | -3.408 | .613 | 30.914 | 1 | .000 | .033 |