Revalidation of the Arizona Risk/Needs Assessment Instrument

Prepared for the:
Arizona Supreme Court
Administrative Office of the Courts
Juvenile Justice Division

Prepared by:
Craig SJ Schwalbe, Ph.D.
Columbia University School of Social Work
1255 Amsterdam Ave
New York, New York, 10027
(212) 851-2228
css2109@columbia.edu

October, 2007

Points of view expressed in this document are those of the author and do not necessarily represent the official position or policies of Administrative Office of the Courts, Arizona Supreme Court.

Table of Contents

Executive Summary and Recommendations …… 2
Introduction…… 4
Introduction to risk assessment…… 5
Arizona Risk/Needs Assessment Instrument……..... 6
Research questions …….. 7
Research Methods ……..... 8
Results ……..... 9
Sample characteristics ..…….. 9
Predictive validity of the risk scale …….. 12
Follow-up analysis: Predictive validity by adjudication status …… 15
Analysis of the needs scale …… 19
Conclusions and Recommendations…….. 22
References …… 24
Appendix …… 26
A.1. Computation of predicted probabilities…… 27
A.2. Source table for Figure 1: 12-month new referral rates……... 29
A.3. Source table for Figure 2: Distribution of risk scores…… 30
A.4. Source table for Figure 3: Distribution of recidivism by risk scores…… 31
A.5. Source table for Figure 4: Needs scale completion rates by risk level.. 32

Summary of Findings and Recommendations.

This study re-validated the Arizona Risk/Needs Assessment Instrument for the Arizona Supreme Court, Administrative Office of the Courts, Juvenile Justice Division. The Arizona Risk/Needs Assessment instrument was developed in the late 1990's to classify risk of recidivism among youths referred to juvenile probation for delinquent offenses and to identify the treatment needs of adjudicated and incorrigible youths (LeCroy, Krysik, & Palumbo, 1998). The Arizona Risk/Needs Assessment Instrument uses overlapping subsets of ten risk factors to calculate a predicted probability of repeat offending for 1st time offenders, 2nd time offenders, and for 3+ time offenders upon referral to the juvenile justice system. Upon adjudication, the Arizona Risk/Needs Assessment Instrument rates potential treatment needs in 13 categories: parenting, health, alcohol and over the counter drug use, illicit drug use, school enrollment status, educational functioning, truancy, behavioral/mental health problems, runaway history, family conflict, assaultive behavior, peer delinquency, and a category of "additional needs." Of interest was the validity of two components of the Arizona Risk/Needs Assessment Instrument: the accuracy of the risk scale to classify risk of recidivism and the potential of the needs scale to improve risk scale classifications.

The study assessed the validity of the Arizona Risk/Needs Assessment Instrument in three cohorts of youths who had a complaint of delinquency during fiscal years 2002, 2003, and 2004. In all, Arizona Risk/Needs Assessment Instruments were gathered for 27,658 youths; the recidivism of these youths was followed until the end of fiscal year 2005. The association between the instrument and 12-month recidivism was tested using descriptive statistics and advanced statistical approaches.

The first main finding of the study is that the predictive validity of the risk scale was stable over the three cohort years and that it was comparable to many other risk assessment instruments used by juvenile justice systems across the country. In general, youths with higher predicted probabilities of recidivism reoffended at higher rates than youths with lower predicted probabilities as would be expected. For instance, youths who had high risk scores, say 81%, reoffended at higher rates than youths who had lower risk scores. Moreover, the predictive validity of the risk scale was similar across gender and race. This finding suggest that the risk assessment instrument continues to perform as expected with respect to its core function – to accurately classify risk of recidivism. However, this positive finding was tempered another: that predicted probabilities calculated by JOLTS using information provided by LeCroy, et al. (1998) do not match observed probabilities for most offenders. Specifically, as the predicted probability of recidivism increases, the gap between predicted probability and actual probability increases. For instance, youths whose predicted probability of recidivism was 81% actually re-offended at a rate of about 50%. This pattern is similar to many other studies which show that, while higher risk scores are consistently associated with higher risk of recidivism relative to lower risk scores (as was the case in this study), the precise probability estimates often do not match observed recidivism rates. As a consequence, most contemporary risk assessment instruments classify risk scores into three or more categories representing "low risk" to "high risk" groups.

The second main finding of this study is that the needs scale cannot at this time be evaluated for its potential to increase the overall predictive validity of the risk assessment instrument. The results showed that over 25% of adjudicated youths were missing needs scales. Trends suggested that higher risk offenders and more serious offenders were more likely to have completed needs scales compared to lower risk and less serious offenders. Because of the high rates of missing needs scales, further analysis into the validity of the needs scale was not completed.

The findings of this study suggest three recommendations:

1. Changes to the risk scale are not recommended. The risk scale is performing as expected and appears to provide useful information to probation officers and to the juvenile courts for disposition and service planning decisions.
2. Further investigation into the interpretation of risk scores by probation officers and judges is recommended to ensure that risk scores are not considered a deterministic probability for individual offenders but rather a relative indicator of youths' propensity to offend compared to youths with smaller risk scores. Subsequent to such an investigation, the department may wish to consider creating a three to five level classification system whereby youths would be grouped into risk groups ranging from "low risk" to "high risk." Such a change would ensure that risk scores were interpreted as relative to other youths rather than as deterministic of individual behavior.
3. Further investigation into the completion rates of needs scales for adjudicated youths is recommended to determine the extent of the problem, its probable causes, and potential remedies. Requiring that probation officers explain incomplete needs scales using a non-intrusive measure modeled after the risk scale would be an intermediate step toward understanding the problem. In addition, other potential barriers to completion should be explored including the availability of information needed to complete the needs scale, the length of the needs scale, and the utility of the needs scale for case planning.

Introduction

Valid risk and needs assessments are key elements of a comprehensive strategy for preventing recidivism in the juvenile justice system (Howell, 1995, 2003). Risk and needs assessment instruments identify high risk/high need offenders so that juvenile courts can target them with sanctions and rehabilitative services to prevent subsequent offending. While the effects of risk and needs assessment on recidivism have not been examined, research on evidence based practices has shown that intervention effects are strongest when targeted toward high risk offenders (Bonta, 1996; Lipsey & Wilson, 1998). Moreover, research documenting the ability of risk assessment instruments to accurately identify high risk offenders is growing (Catchpole & Gretton, 2003; Funk, 1999; Gretton, 2001; Hoge, 2002; Ilacqua, Coulson, Lombardo, & Nutbrown, 1999; Krysik, 2002; Schmidt, Hoge, & Gomez, 2005; Schwalbe, in-press; Schwalbe, Fraser, & Day, 2007; Schwalbe, Fraser, Day, & Arnold, 2004; Schwalbe, Fraser, Day, & Cooley, 2006). Buoyed by this promise, utilization of risk and needs assessment instruments has grown from 33% of state juvenile justice systems in 1990 to 86% by 2003 (Griffin & Bozynski, 2003; Towberman, 1992).

Arizona is a leader in juvenile justice risk and needs assessment. The first risk assessment instrument was implemented in 1990 and was substantially revised in 1994. This revision was followed by a large scale study in 1998 by LeCroy & Milligan Associates, Inc. to evaluate the 1994 risk assessment instrument and to develop a replacement. The replacement risk scale was more accurate than the 1994 instrument and was subsequently adopted in 2000. Development of a companion needs scale commenced and was adopted in 2002. Together, the Arizona Risk/Needs Assessment Instrument provide the assessment platform upon which case processing decisions and service planning decisions are made. The risk scale is completed for all youth upon referral to the juvenile justice system for a delinquent offense. The JOLTS system calculates a predicted probability risk score for individual youths based on probation officer ratings of ten risk factors. The need scale is completed for all youths adjudicated delinquent by the juvenile court. The need scale is comprised of a series of questions about risk and need factors in 13 categories that may indicate the need for interventions. Needs scale categories include: parenting, health, alcohol and over the counter drug use, illicit drug use, school enrollment status, educational functioning, truancy, behavioral/mental health problems, runaway history, family conflict, assaultive behavior, peer delinquency, and a category of “additional needs.”

The purpose of the present study is two-fold. First, the study aims to revalidate the Arizona Risk/Needs Assessment Instrument. Periodic validation studies are recommended to ensure that risk assessment instruments capacity to classify risk of recidivism is sustained over time. Second, the study aims to determine whether or not the needs assessment instrument could be utilized to further strengthen risk assessment predictive validity.

Introduction to Risk Assessment

The purpose of risk assessment in juvenile justice is to classify risk of future offending. Risk assessment instruments do this by measuring diverse risk factors for recidivism. Two extensive empirical reviews inform this effort. Lipsey and Derzon (1998) reviewed 34 longitudinal studies of the correlates of serious criminal or violent behavior, defined as any index offense, while Cottle, Lee and Heilbrun (2001) focused on 22 studies of the risk factors associated with re-arrest, re-adjudication, probation violation, and recommitment. Both studies found risk factors with strong effects in the domains of offense history, psychological factors, family factors, and peer factors. Most risk assessment instruments incorporate risk factors from two or more of these categories.

In most risk assessment instruments, scores from risk factors are added together to derive a "cumulative risk score" which is re-classified into three or more risk classes (i.e. low-risk, medium-risk, high-risk). These classifications correspond to an array of graduated sanctions and court interventions designed to prevent recidivism (Howell, 1995, 2003). Risk assessment instruments have high levels of predictive validity when juveniles are classified into risk-classes with widely varying rates of recidivism. In addition, well performing risk assessment instruments classify relatively more youths in the lower and higher risk categories and relatively fewer youths in the medium risk categories. This is because medium risk groups tend to have recidivism rates that approach the sample base rate. In effect, the risk assessment instrument adds little benefit for youths classified as "medium risk."

Risk assessment instruments should correctly classify offenders across diverse demographic groups. For example, a "high risk" classification should convey a similar meaning with respect to the probability of recidivism for males, females, Hispanic juveniles, and white juveniles, among others. Under ideal conditions, group differences in rates of re-offending should be accounted for by differences in the proportions of juveniles classified into lower and higher risk categories. In statistical modeling terms, we would expect that the effects of race/ethnicity or gender on recidivism would be mediated by cumulative risk level.

The gold standard for risk assessment development is the actuarial approach. The hallmark of the actuarial approach is an empirical development strategy that separates instrument development from validation (Gottfredson & Tonry, 1987; Shlonsky & Wagner, 2005). The first step in risk assessment construction is to identify risk factors to include in an instrument. This process is completed with an "estimation sample" of juveniles in which risk factors that predict recidivism at a statistically significant level are identified and gathered together into an index. The second step in risk assessment construction is to test its "predictive validity" or accuracy. This is done with a separate "validation" sample of juveniles in which the overall predictive validity of the risk assessment instrument is tested and reported. In large sample studies, research participants can be randomly assigned into estimation and validation samples. In other cases, validation can follow risk assessment development using an independent sample drawn from the same population. The distinction between estimation and validation is the strength of the actuarial approach. In general, predictive validity estimates for

estimation samples tend to be inflated because they "over-fit" the data whereas predictive validity estimates for validation samples tend to be more realistic. But by separating the validation process from the estimation process, the actuarial approach results in predictive validity estimates that should be more stable over time.

The Arizona Risk/Needs Assessment Instrument

LeCroy et al. (1998) is a landmark study in the history of risk assessment in Arizona. The purpose of their study was to evaluate the risk assessment instrument developed in 1994 and to develop a new risk assessment as a replacement. LeCroy et al. is set apart from other risk assessment studies by its systematic attention to the reliability and utility of risk assessment measures in addition to its focus on accuracy, or predictive validity. Moreover, the study is unique by its use of a large sample and by its fidelity to the actuarial approach.

LeCroy et al. (1998) found that the 1994 risk assessment instrument had multiple problems in practice. Chief among these were low completion rates. Across counties, completion rates of the 1994 risk assessment instrument ranged from 13% to 76%. Surveys with probation officers indicated that up to 84% believed that their subjective rating of risk was more accurate than the 1994 risk assessment instrument. Indeed, LeCroy et al.'s analysis confirmed probation officers' perceptions. Compared to probation officers, the predecessor risk assessment classified more youths into the uninformative "medium risk" group (74%). And, recidivism rates for youths classified as high risk with the predecessor risk assessment instrument were lower than youths classified as high risk by subjective judgement (.50 vs. .68, respectively).

To develop a revised risk assessment instrument, LeCroy et al. (1998) relied on large samples of first time (N=11,755), second time (N=5,921) and third-plus time offenders (N=6,584). Each group was randomly divided into estimation (60%) and validation (40%) samples. Candidate risk factors were coded by probation officers as a part of routine practice and entered into JOLTS, the statewide management information system. Multivariate logistic regression equations were used to identify statistically significant predictors and to weight individual items. This process resulted in risk assessment instruments of five items (1st and 2nd time offenders) or six items (3rd+ time offenders). Regression weights were used to estimate predictive probabilities for the validation sample which were in turn classified into three offender groups: low risk, medium risk, and high risk. Recidivism rates ranged from 24% for low risk offenders to 82% for the high risk offenders. Moreover, relatively few youths were classified into the medium risk group (20% - 44%). Sensitivity analysis by race and gender found no differences in predictive validity across these groups. Inter-rater reliability of the revised risk assessment instrument was tested using a sample of 23 probation officers who rated six cases based on file information. Among probation officers, percent agreement for each of the risk factors ranged from 76% to 96%.

The risk scale of the Risk/Needs Assessment Instrument is essentially a composite of three risk assessment instruments – one each for 1st time offenders, 2nd time offenders, and 3rd+ time offenders. Predicted probability risk scores are calculated for each group

of offenders from a subset of ten risk factors. Three risk factors are shared across the instruments: juvenile's relationship with his/her family involves frequent/intense conflict or is alienated/assaultive; juvenile used, or is suspected of using drugs within the past year; and juvenile ever truant or extensive absenteeism from school. The 1st time offender instrument adds two risk factors: current offense is a status offense, juvenile has ever been assaultive. The 2nd time offender instrument adds two risk factors: juvenile not currently enrolled in a public, private, or home school on a regular basis; and juvenile has behavior problems/mental health issues. The 3+ time offender instrument adds three risk factors: friends involved, or suspected to be involved, in delinquency; number of prior complaints; and runaway, runaway attempts, known or suspected. Each item is dichotomously scored (yes/no). In all cases but 'school enrollment,' a "yes" response indicates the presence of risk. Each item is weighted using logistic regression parameter estimates developed by LeCroy et al. (1998) and predicted probability risk scores are calculated using a standard formula. For instance, a 1st time offender with two risk factors – family conflict and drug use – has a 50% probability of recidivism according to the LeCroy et al. method. A 2nd time offender with the same two risks has a 64% probability of recidivism and a 3rd+ time offender has a 72% probability of recidivism. The computation procedures are shown in the appendix.

The results of LeCroy et al. (1998) were compared to the results of other risk assessment studies in a meta-analysis completed by Schwalbe (*in-press*). A meta-analysis uses statistical tests to compare the results from a group of similar studies and to identify the average effects across studies. In this case, Schwalbe was interested in identifying the average level of risk assessment predictive validity across many risk assessment instruments and was interested in identifying characteristics of risk assessment instruments that characterized "high performing" instruments. Searching for published and unpublished studies of juvenile risk assessment instruments, Schwalbe identified 28 studies that evaluated 28 risk assessment instruments. Across these instruments average levels of predictive validity were moderate (*AUC*=.640)[1] but consistent with those found in risk assessment studies with adults in the criminal justice system. The revised Arizona Risk/Needs assessment instrument for first time offenders compared favorably to this average (AUC=.716) while the second time and third time offender instruments performed at the average level (AUC=.641).

Research Questions

The proposed study was motivated by three questions:

1. To determine the stability of the Arizona risk assessment instrument predictive validity by comparing predictive validity estimates in a new sample to estimates originally shown by LeCroy, Krysic, and Palumbo (1998).
2. To identify items from the Needs assessment instrument that predicts recidivism.
3. To increase the predictive validity of the Arizona juvenile risk assessment instrument by adding items from the comprehensive needs assessment.

[1] AUC is a statistical measure of predictive validity. It ranges from .500 to 1.0; higher values represent higher levels of predictive validity, values closer to .500 indicate predictive validity approaching chance, i.e. coin toss.

Research Method

The study tracked the recidivism of three cohorts of youths referred to the juvenile justice system. All data for the study were abstracted from the JOLTS data-base. Measures included:

1. Demographic information about the youths.
2. Offending characteristics.
3. Record of the intake decision (i.e. dismissal, consequence, petition, adjudication)
4. Risk/Need Assessment Instruments for each youth completed at intake in the case of the risk scale and after adjudication in the case of the needs scale.
5. Reason for non-completion for youths whose Risk/Need Assessment Instrument was incomplete.
6. Recidivism defined as a new referral to the juvenile justice system on or before June 30, 2005. Referrals to the court for technical violation of probation were excluded.
7. Placement in court-funded out-of-home placements.
8. Violation of probation.

The sample for this study was a random selection of youths referred to any juvenile probation department in Arizona during three consecutive fiscal years: FY02 (July 1, 2001 – June 30, 2002), FY03 (July 1, 2002 – June 30, 2003), and FY04 (July 1, 2003 – June 30, 2004). It included juveniles who were eventually referred to the juvenile court as well as those who were diverted following the initial assessment. Juveniles who were referred multiple times during the year were eligible for inclusion based on their first referral within a given year. A separate random sample of youths was selected during each fiscal year resulting in three distinct cohorts.

The target sample size for each cohort was 12,000 youths; in total, records for 36,139 youths were selected. However, cases were omitted from the analysis for three reasons. 8,034 youths were assessed after their 17th birthday. These youths were omitted from the analysis to provide all youths with a 12-month follow-up period. 131 youths were under 8 years old at the time of assessment. According to Arizona state juvenile justice policy, youths under 8 years old are ineligible to be charged with delinquent offenses. Finally, 330 juveniles were charged with “administrative” offenses. This offense category is used to classify non-delinquency-related court matters for which a court file is nevertheless created.

The sampling strategy resulted in an overlap among the cohorts. 34.4% (n=3,229) of youths in FY02 were also selected for FY03 and 33.3% (n=3,048) of youths in FY03 were also selected for FY04. The structure of the data permitted identifying which youths were selected for multiple cohorts but did not permit identifying individual juvenile records across the cohorts. That is, while it was possible to determine that a specific FY02 juvenile was also selected for FY03, changes to the case identifier made it impossible to link records across cohort years. Therefore, each cohort was treated as a unique sample in the analysis. The final sample sizes for this study were N_{FY02}= 9,398; N_{FY03}= 9,153; and N_{FY04}= 9,107.

Results

Sample Characteristics

Table 1
Demographic characteristics

		FY02 (n=9.398)	FY03 (n=9,153)	FY04 (n=9,107)
Male (%)		64.3	64.9	65.5
Race/Ethnicity (%)	White	47.5	48.1	44.8
	Latino	38.6	38.4	40.9
	Black	6.5	6.7	7.1
	Native American	5.8	5.4	5.9
Age (Mean, S.D.)		14.7 (1.7)	14.8 (1.7)	14.7 (1.7)

Table 1 shows the demographic characteristics of the three cohorts. Patterns were similar across each cohort. The majority of youths were male. Nearly half were white; over a third were Latino. The average youth in each cohort was about 15 years old at the time of their initial assessment.

Table 2
Offending characteristics

	FY02 (n=9.398)	FY03 (n=9,153)	FY04 (n=9,107)
Offense Severity (%)			
Felony, person	3.7	3.8	4.1
Felony, property	8.5	8.6	8.5
Obstruction of justice	7.1	7.1	6.5
Misdemeanor, person	9.7	10.3	10.4
Drug offenses	9.2	9.6	10.0
Public peace	21.1	22.2	21.3
Misdemeanor, property	17.4	17.3	17.1
Status offense	23.3	21.1	22.2
Offending History			
First time offender (%)	59.9	61.1	60.8
Number prior offenses (Mean, S.D.)	1.3 (2.8)	1.3 (2.7)	1.2 (2.5)
Age at first referral (Mean, S.D.)	13.3 (2.0)	13.4 (2.0)	13.4 (2.0)
Violation of Probation (%)	6.7	6.8	7.4

Table 2 shows the offending characteristics of the three samples. It shows that around 60% of youths were first time offenders and that the majority had minor offenses (i.e. public peace, status offenses, and misdemeanor property). For each cohort, the number of prior offenses ranged widely from 0 to 44 for FY02 youths, 0 to 48 for FY03 youths,

and 0 to 32 for FY04 youths. 95% of youths had six or fewer prior offenses. The average age of youths at first referral was 13.3 years old for all cohorts. 30.7% of FY02 youths were 12 years old or younger at their first referral; 29.4% of FY03 and 29.7% of FY04 youths were 12 years old or younger at their first referral. Around 7% of each cohort were referred for a violation of probation (offense code 1087, 1090, 1091). It should be noted that additional analysis (not shown) indicated that the predictive validity of the risk scale was not altered with the inclusion of VOP youths.

Table 3
12-month recidivism rates by gender and race

	FY02	FY03	FY04
Full sample	33.0	32.1	30.5
Gender			
Males	35.3	34.5	33.1
Females	30.0	27.7	25.5
Race			
White	31.3	30.4	28.4
Latino	35.4	34.9	32.5
Black	38.8	33.4	34.8
Native American	28.1	29.0	29.7

Table 3 and Figure 1 show 12-month recidivism rates for the three cohorts[2]. Overall, 12-month recidivism rates ranged from 31-33%. In all years, males had slightly higher recidivism rates than females though patterns in recidivism rates by race were less stable. African American youths had the highest recidivism rates in two years; Native American youths had the lowest recidivism rates in two years. Patterns of recidivism over time are indistinguishable for the three cohorts. By nine months, 27-28% of juveniles reoffended; by 12-months, 31-33% of juveniles reoffended. Across the total length of the study, 47.6% of FY02 youths, who were followed for up to 4 years, reoffended; 42.5% of FY03 youths, who were followed for up to 3 years, reoffended; and 34.4% of FY04 youths, who were followed for up to 2 years, reoffended[3]. These patterns suggest that the first year after intake represents a high risk year for most offenders. Indeed, of FY02 youths who eventually reoffended, 59% reoffended in the first nine months after intake and 69% reoffended in the first year after intake.

[2] The source table for this and all figures in the report can be found in the appendix.
[3] Differences in overall rates can be accounted for by the varying lengths of follow-up for each group. Recidivism for FY02 youths were tracked from 3 to 4 years; recidivism for FY03 youths was tracked from 2-3 years; recidivism for FY04 youths was tracked for 1 to 2 years.

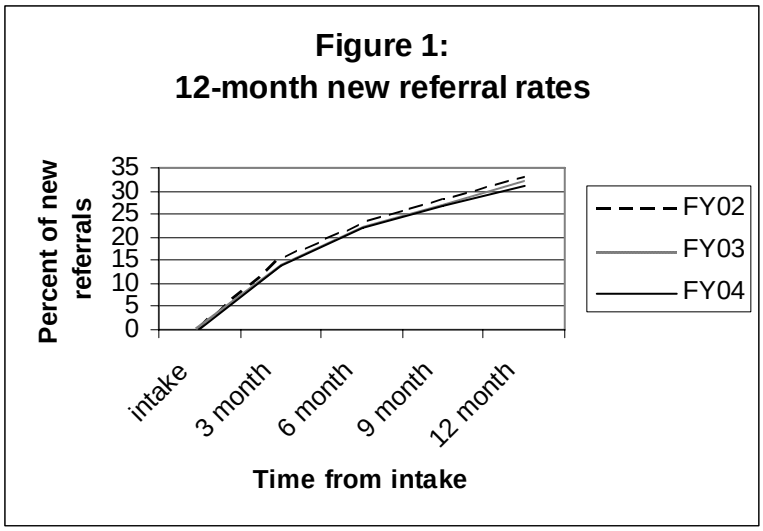
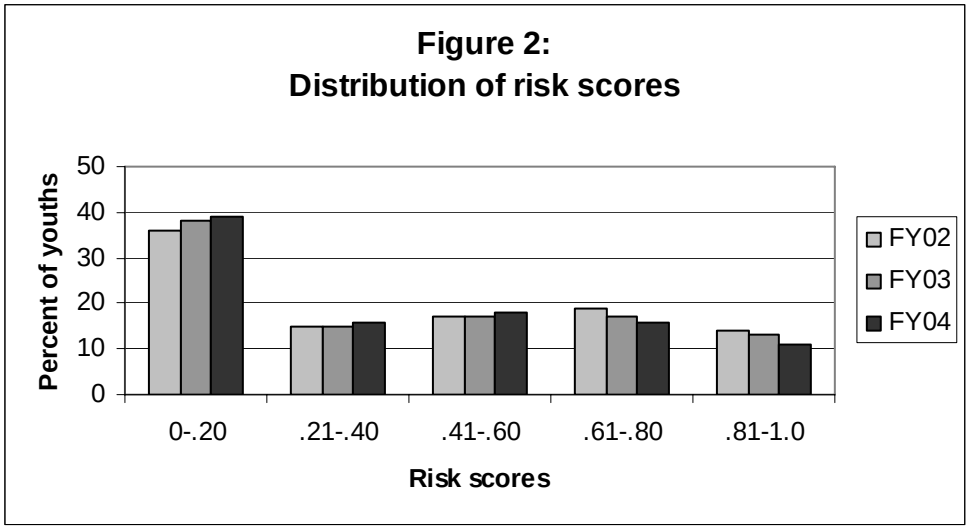


Table 4 shows the distribution of risk factors from the risk scale for the three cohorts. As with other sample characteristics, the cohorts are more similar than different. By far, the most prevalent risk factor is peer delinquency. Prevalence of peer delinquency problems approaches 70% for each cohort. The next most prevalent risk factors are truancy ,drug use, assaultive behavior, and behavioral/mental health problems. Problems with school enrollment and running away are the least prevalent risk factors.

Table 4
Prevalence of risk factors

	FY02 (n=9.398)	FY03 (n=9,153)	FY04 (n=9,107)
Family relationship (%)	29.6	27.6	24.2
Assaultive (%)	39.2	37.7	36.5
Drug use (%)	41.8	40.4	37.9
Truancy (%)	46.6	41.7	41.9
Not enrolled in public, private, or home school (%)	16.2	17.7	18.2
Behavior/mental health (%)	31.3	28.2	26.5
Peer delinquency (%)	69.4	69.1	66.6
Runaway (%)	22.8	19.2	17.5
Status offense (%)	30.3	26.0	27.9
Number prior complaints (mean, s.d.)	1.3 (2.8)	1.3 (2.7)	1.2 (2.5)

The distribution of risk scores from the risk scale was analyzed next. Total risk scores range from .19 for lowest risk youths to .91 for highest risk youths. The average risk score across cohorts ranges from .50 to .53. Figure 2 presents the distribution of risk scores graphically. For this analysis, risk scores were divided into five risk categories: 0-.20, .21-.40, .41-.60, .61-.80, .81-1.0. About 37% of youths in each cohort have risk scores in the lowest range. The remaining youths are evenly distributed across the risk categories.



Predictive validity of the risk scale

The analysis next turned to the risk scale predictive validity. First, the completion rate and reason for non-completion was examined. As shown by LeCroy et al. (1998), completion rates for the 1994 risk assessment instrument were just 62%, lower than desired. In the present study sample, a large majority of youths who were referred to the juvenile justice system had complete risk scales – 87.5% across all cohorts. Probation officers recorded justifications for all youths with incomplete risk assessments. The most prevalent reason for non-completion was "dismissed, no petition filed" (30.7%) followed by "unable to locate" (30.0%) and by "matter handled by lower jurisdiction (13.2%).

Table 5
Correlation coefficients for the association of risk factors with 12-month recidivism

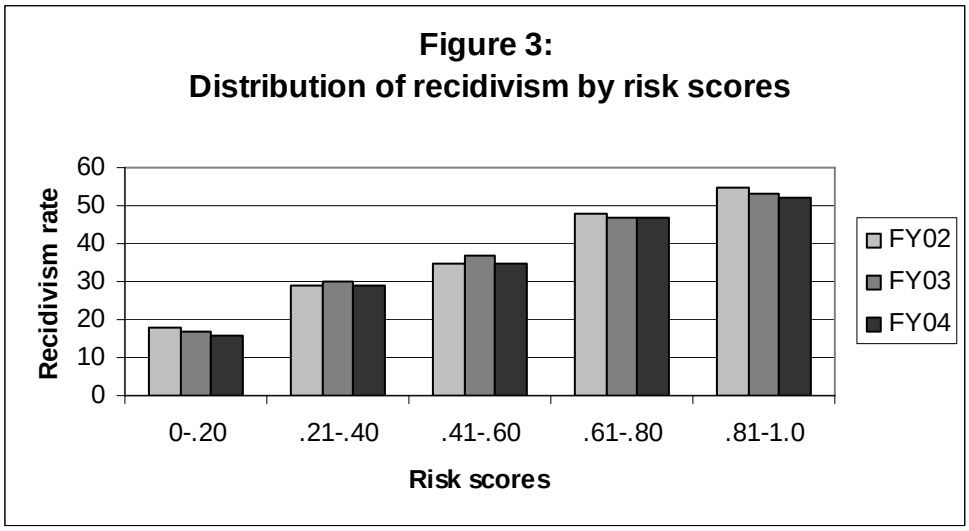
	FY02	FY03	FY04
Family relationship	.25	.26	.23
Assaultive	.19	.20	.18
Drug use	.25	.20	.26
Truancy	.18	.21	.20
Not enrolled in public, private, or home school	.11	.07	.09
Behavior/mental health	.19	.19	.19
Peer delinquency	.18	.14	.18
Runaway	.24	.23	.23
Status offense	.10	.15	.12
Prior complaints	.15	.13	.16

Next, the association between individual risk factors with recidivism was analyzed. Table 5 shows the association between individual risk factors and 12-month recidivism rates using correlation coefficients[4]. As with previous analyses, patterns were consistent

[4] The table shows tetrachoric correlation coefficients. Like all correlation coefficients, tetrachoric correlation coefficents describe the strength of a relationship between two variables on a scale from -1 to

across cohort years. In general, all risk factors predicted recidivism as expected. Risk factors with relatively stronger predictive validity (r>.20) across all years include family relationships, drug use during the past year, and runaway attempts during the past year. Truancy had relatively stronger predictive validity in two years and assaultive behavior had relatively stronger predictive validity in one year. School enrollment status and status offences had the weakest association with recidivism in all years.

Finally, the predictive validity of the risk scale was examined. Figure 3 shows the 12-month predictive validity of the risk assessment instrument with descriptive statistics. As in the previous analysis, risk scores were divided into five risk categories. Figure 3 shows that recidivism rates increase with increases in risk as expected. Recidivism rates range from 16-18% for the lowest risk group to 52-55% for the highest risk group. However, the graph also shows that recidivism rates for three out of five risk classes fall outside of the predicted range. Fewer youths from the three highest risk classes re-offended than was predicted by the risk assessment instrument.



While the descriptive analysis presented in Figure 2 illustrates important trends in the association between the risk score and recidivism, statistical analyses are required to describe predictive validity in a way that makes it comparable to other studies. Statistical analyses have the advantage of describing predictive validity with a single number that controls for variation in key sample characteristics like recidivism base rates and distribution of risk factors. Two statistical tests were employed in this study: Area Under the Curve (AUC) and point-biserial correlation coefficients. The AUC is used to describe the predictive validity of many diagnostic and screening tests. It ranges from .500 to 1.0 and is interpreted as the probability with which a randomly selected repeat offender will have a higher risk score than a randomly selected non-repeat

---

+1. Correlation coefficients that approach zero indicates weak relationships, correlation coefficients approaching -1 indicates strong inverse relationships, and correlation coefficients approaching +1 indicates strong positive relationships. Correlation coefficients of “moderate” strength (e.g. r=.30) are common in social science research.

offender. Risk assessment instruments that have high AUC scores will tend to concentrate repeat offenders among the higher risk scores. Risk assessment instruments with AUC scores approaching .500 classify recidivism at the level of chance (i.e. coin toss). For example, at AUC=.550, randomly selected repeat offenders will have higher risk scores than randomly selected non-repeat offenders just 55% of the time while at AUC=.75, randomly selected repeat offenders will have higher risk scores than randomly selected non-repeat offenders 75% of the time. The point biserial correlation describes the strength of the association between risk score and recidivism on a scale from -1 to +1. Scores that approach zero describe a weak relationship between variables; scores that approach -1 describe a progressively stronger inverse relationship between variables; scores that approach +1 describe a progressively stronger positive relationship between variables. Rice and Harris (1995) have shown that the AUC is more stable than the point biserial correlation coefficient when describing risk assessment predictive validity. For this reason, the AUC is gaining increased acceptance as the statistical test of choice for describing risk assessment predictive validity. Nevertheless, correlation coefficients are more widely known and understood. Therefore, both are reported in this study.

Table 6
12-month predictive validity of the risk scale: AUC scores

	FY02	FY03	FY04
Full sample	.652	.652	.659
Gender			
Males	.648	.655	.652
Females	.661	.644	.670
Race			
White	.652	.659	.670
Latino	.655	.644	.644
Black	.641	.626	.631
Native American	.614	.663	.629

Table 6 reports AUC scores for the risk scale and Table 7 reports point biserial correlation coefficients for the risk scale. Predictive validity measures in both tables show the same pattern: highly stable predictive validity scores across the three cohorts for the full sample, for males, for females, and for White and Latino youths. Scores are more variable across cohorts for Black and Native American youths. However, this can be attributed to their relatively small sample size. Moreover, predictive validity estimates for these groups along with all others compares favorably to the average predictive validity across all risk assessment instruments shown by Schwalbe (in-press; average AUC=.640, average point biserial correlation coefficient=.25). Predictive validity estimates for the present study are lower than those calculated for LeCroy et al. (1998; 1st time offender, AUC=.716). However, a certain amount of “shrinkage” is to be expected when validating risk assessment instruments on new samples (Silver, Smith, & Banks, 2000). Overall, these findings suggest that the Arizona Risk/Need Assessment

instrument predicts recidivism as well or better than the average risk assessment instrument used in the field and that these findings are stable over time[5].

Table 7
12-month predictive validity of the Risk Scale: Correlation coefficients

	FY02	FY03	FY04
Full sample	.25	.25	.26
Gender			
Males	.25	.26	.25
Females	.26	.23	.26
Race			
White	.25	.26	.27
Latino	.26	.24	.24
Black	.24	.21	.22
Native American	.18	.26	.21

Follow-up analysis: Predictive validity by adjudication status
An additional analysis was undertaken to compare the predictive validity of the risk scale in the total sample with sub-samples of adjudicated youths (N_{FY02}=2.357, N_{FY03}=2,175, N_{FY04}=2,127). The literature on risk assessment tends to include higher risk samples of adjudicated youth or institutionalized youths. This analysis will allow more direct comparisons to these studies.

As would be expected, adjudicated youth tend to be older than non-adjudicated youths (15.1 years vs. 14.6 years old, t(28,00)=21.45, p<.0001), have higher risk scores (.65 vs. .46, t(28,000)=55.2, p<.0001), and more serious offenses. More males than females were adjudicated (28% vs. 17%) while differences in adjudication rates across race were small (white=22.5%, latino=25.6%, African American=27%, Native American=26%). Over all cohorts, average recidivism rates were higher for adjudicated youths than non-adjudicated youths (42.3% vs. 28.6%). Risk assessment completion rates were higher for adjudicated youths than non-adjudicated youths (96% vs. 85%).

[5] During the course of this study, a potential JOLTS programming error was discovered in the calculation of risk scores for 3+ time offenders. It was discovered that risk scores were adjusted for youths with five or more referrals rather than four or more referrals as should have been the case. Preliminary analysis with the present study sample suggested that the error had no impact on statistical measures of predictive validity. Moreover, whereas the average youth with four referrals should have had a risk score of .82, they instead received a risk score of .77, that is, their risk scores were underestimated slightly. To the extent that higher risk scores are related to more severe sanctions, youths were not disadvantaged as a consequence of this error.

Table 8
12-month predictive validity of the risk scale among adjudicated youths: AUC scores

	FY02 (N=2,357)	FY03 (N=2,175)	FY04 (N=2,127)
Full sample	.597	.579	.603
Gender			
Males	.585	.598	.621
Females	.629	.540	.577
Race			
White	.604	.580	.623
Latino	.585	.586	.591
Black	.568	.568	.558
Native American	.559	.547	.610

Table 8 and table 9 compare the predictive validity of the risk scale with adjudicated youths for the full sample of adjudicated youths and across gender and race. Patterns are similar in both tables. In general, predictive validity scores are lower for adjudicated youths than for the full sample reported above. Variation in scores across gender and racial categories also increased although sampling error resulting from small sample sizes is the likely cause. For example, Native American youths number just 155 (FY02), 117 (FY03), and 139 (FY04).

Table 9
12-month predictive validity of the risk scale among adjudicated youths: Correlation coefficients

	FY02 (N=2,357)	FY03 (N=2,175)	FY04 (N=2,127)
Full sample	.17	.14	.18
Gender			
Males	.15	.17	.21
Females	.22	.07	.13
Race			
White	.18	.14	.21
Latino	.15	.15	.16
Black	.12	.12	.10
Native American	.10	.08	.19

The next analysis sought to explain the reduction in predictive validity for adjudicated youths. One explanation is that there is less variation in risk scores among adjudicated youths compared to non-adjudicated youths. This would tend to suppress predictive validity estimates in statistical tests. However, the AUC is less susceptible to this type of bias than traditional correlation coefficients; results reported above showed that both AUC scores and correlation coefficients are smaller for adjudicated youths compared to non-adjudicated youths. An alternative explanation is that adjudicated youths are

treated differently than non-adjudicated youths and this treatment alters the relationship between risk scores and recidivism.

Two variables were available to test this hypothesis: violation of probation and length of time in out-of-home placement. Adjudicated youths were 8.8 times more likely to be placed in out of home care than non-adjudicated youths (4.7% vs. .56%). Theoretically, out-of-home placements reduce the opportunity for delinquent behaviors for the period of placement and may have treatment-related effects that endure post-placement. Also, adjudicated youths by definition are at risk of VOP whereas non-adjudicated youths are not. Among adjudicated youths, 19% had at least one VOP within 12-months. Theoretically, VOP represents an aversive consequence for pre-delinquent behavior which should deter future involvement in delinquency. Both VOP and out-of-home placements should have the strongest effects with higher risk youths.

Logistic regression analysis was used to identify the joint effects of the risk assessment instrument, gender, race, VOP, and out-of-home placement on 12-month recidivism rates. Logistic regression is a multivariate procedure to estimate the effects of a set of predictor variables on the probability of an event like recidivism. The results yield odds ratios (OR) for each predictor variable. If OR approaches one (i.e. OR=1), this indicates that changes in the predictor variable are associated with no changes to the probability of recidivism. OR<1 indicates that any 1-unit increase in the predictor variable is associated with a decrease in the probability of recidivism; OR>1 indicates that any 1-unit increase in the predictor variable is associated with an increase in the probability of recidivism.

Table 10 and table 11 show the results of the analysis. Table 10 includes race, gender, and the risk score in the analysis. Results for race are inconsistent across cohort years suggesting that recidivism rates may not vary according to race when gender and the risk score are controlled. Results for gender suggest that when controlling for race and risk score, males are 1.4-1.5 times more likely to reoffend than females. Said another way, the odds of recidivism are 40-50% higher for males than for females. Results for the risk score indicate that for every .10 increase in the risk score (i.e. from .20 to .30), the odds of recidivism increase by a factor of 1.2 or by 14-19%.

Table 10
Logistic regression of race, gender, and risk score on recidivism

	Odds Ratio		
	FY02	FY03	FY04
Latino	1.3	*	1.3
Black	2.0	*	*
Native American	*	*	*
Other race	*	.3	*
Male	1.5	1.4	1.4
Risk score	1.16	1.14	1.19

*Note. Asterisks indicate non-significant results

Table 11 shows the same statistical models when VOP and out-of-home placements are included. Of interest in these analysis is the change in the OR for risk score. If VOP and out-of-home placements reduce recidivism rates for high risk youths, we would expect that the risk score OR would increase. The results show that VOP did not reduced recidivism as expected but that length of time in out-of-home placement did reduce recidivism during the 12-month follow-up period – for every month spent in out-of-home placement, the odds of recidivism decreased by about 20%. However, contrary to the hypothesis, neither VOP nor out-of-home placements explained the reduction in the predictive validity of the risk score for adjudicated youths. That is, the magnitude of the risk score odds ratio did not change when VOP and out-of-home placement was included in the analysis.

Table 11
Logistic regression of race, gender, risk score, VOP, and placement on recidivism

	Odds Ratio		
	FY02	FY03	FY04
Latino	1.3	*	1.2
Black	2.0	*	*
Native American	*	*	*
Other race	*	.3	*
Male	1.5	1.4	1.4
Risk score	1.17	1.16	1.20
VOP	*	*	*
Months placed	.80	.78	.78

*Note. asterisks indicate non-significant results

Analysis of the Needs Scale

The analysis of the needs scale began with an examination of its completion rate and trends associated with complete versus incomplete needs scales. Table 12 presents the number of completed needs scales for the full samples in each cohort year. Completion rates for FY02 were small (4%) indicating that the needs scale had not been fully implemented during that fiscal year. Completion rates for FY03 and FY04 were similar (28%). Further analysis focused on these two cohorts.

Table 12
Needs scale completion rates by cohort year

Cohort Year	Completed needs assessments
FY02 (N=9,398)	404 (4.3%)
FY03 (N=9,153)	2,600 (28.4%)
FY04 (N=9,107)	2,606 (28.6%)

The needs scale is mandated for youths who are adjudicated by the juvenile court. Therefore, one would expect that completion rates would differ according to referral status. Table 13 presents needs scale completion rates by referral status. Completion rates were lower for youths who were not petitioned to the juvenile court. For instance, about 19-20% of youths who received a consequence by the probation officer with no further petition to the juvenile court had a completed Needs Scale on record. Completion rates were higher for youths who were adjudicated (FY03=63.5%, FY04=68.9%). Although completion rates increased over the two years, they are lower than what might be expected given the policy mandate to complete needs scales for all adjudicated youths.

Table 13
Needs scale completion rates by referral status

	Cohort 03		Cohort 04	
	n	% completed	n	% completed
No consequence – no petition	1,947	9.0	1,922	9.1
Consequence, no petition	4.056	20.1	3,942	18.6
Consequence, petition	422	37.2	449	33.0
Petition directly	2,178	53.2	2,787	55.5
Adjudicated	2.175	63.5	2.127	68.9

The next stage of the analysis explored potential explanations for the observed completion rates among adjudicated youths. If completion is randomly distributed across the adjudicated youths, then further exploration of the needs scale can be justified. However, if completion is a non-random process, the association between items on the completed needs scale and recidivism would be inaccurate and would potentially diminish the predictive validity of the risk scale. Completion patterns were explored for offense severity, final disposition, and risk level.

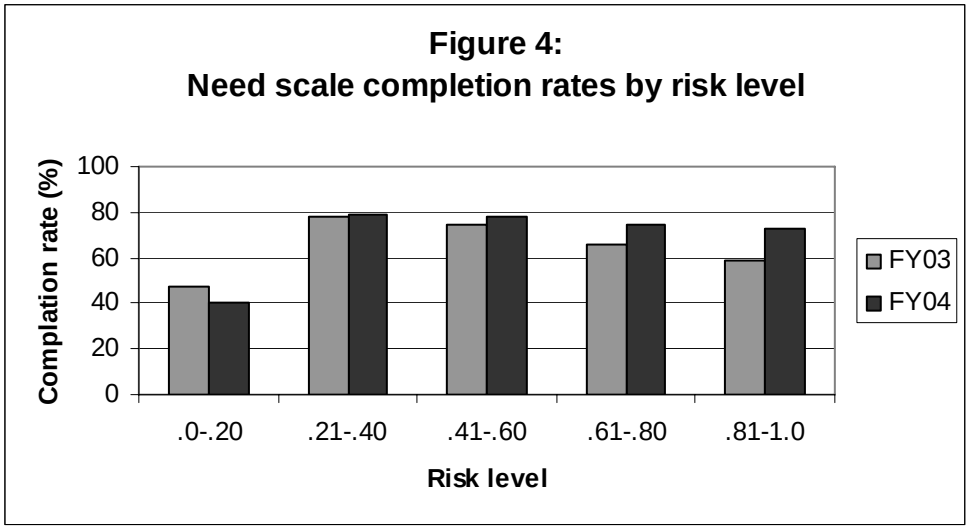
Table 14 shows completion rates by offense severity. It shows that completion rates increased from FY03 to FY04 for most categories. In both years, completion rates were highest for felony-person offenses followed by drug offenses and felony-property offenses. Table 15 shows completion rates by disposition type. By far, probation was the most common disposition ordered across both years (77-78%). Completion rates for these youths increased from 68% in FY03 to 72% in FY04. Likewise, completion rates increased for the small numbers of youths who were placed on JIPS, who were committed to ADJC, and who were ordered to protective supervision. Finally, Figure 4 shows the distribution of completion rates across risk scores. At the lowest risk level, completion rates were low (around 40%) in both cohorts. In FY03, completion rates increased markedly for the second risk level to nearly 80% and declined slightly for the third, fourth, and fifth risk levels to a low of approximately 60%. In FY04, completion rates increased to 80% for the second risk level and maintained this increase for the third, fourth, and fifth risk levels. Together, these findings show that completion rates increased from FY03 to FY04. Across both years, completion rates were higher for more serious felony-type offenses and did not vary markedly by disposition type. On the other hand, whereas high risk offenders were more likely to have a completed needs scale, low risk offenders were less likely to have a completed needs scale.

Table 14
Needs scale completion rates by offense severity

	Cohort 03 (n=2,175)		Cohort 04 (n=2,127)	
	n	% completed	n	% completed
Felony, person	161	70.8	185	78.9
Felony, property	419	67.3	435	72.2
Obstruction of justice	467	50.1	410	61.0
Misdemeanor, person	222	65.8	212	74.1
Drug offenses	320	72.2	294	76.2
Public peace	292	69.5	300	62.3
Misdemeanor, property	192	56.3	192	66.7
Status offense	102	61.8	99	59.6

Table 15
Completion rates by disposition type

Disposition type	FY03		FY04	
	N	Completion rate (%)	N	Completion rate (%)
Probation	1708	67.6	1,637	71.6
JIPS	183	55.7	221	69.2
Protective supervision	81	46.9	70	60.0
ADJC commitment	23	30.4	17	70.6
Other	180	44.4	182	47.3



Despite the trend toward greater completion rates, large numbers of incomplete need scales preclude a meaningful analysis of the items within the need scale. The number of missing need scales exceed the missing data threshold (10%) considered to be acceptable in the LeCroy et al. (1998) study. To be sure, the validity of any analysis of the need scale, whether to describe the prevalence of need within the population of adjudicated youths or to assess the predictive validity of individual need scale items, would be threatened by the high likelihood that the completed need scales represent a non-random subset of all adjudicated youths. For this reason, further analysis of the need scale items was not conducted and is not recommended at this time.

Conclusions and Recommendations

There are four conclusions that can be drawn from this study. First, this study demonstrated that the predictive validity of the Arizona Risk/Needs Assessment Instrument compares well against its peers and is stable over time. The scoring error discovered during the course of this study (see footnote 5) appears to have a negligible effect on predictive validity. Moreover, analysis by race and gender showed that the instrument performs similarly for these groups. In general, the predictive validity findings of the present study mirror the results of LeCroy et al. (1998) and suggest that the juvenile courts are well served by their current risk assessment instrument.

Second, this study was consistent with others showing that predicted probabilities often do not correspond to observed offending rates (Dean & Duggan, 1968). In the present study, actual recidivism rates did not match predicted probabilities derived from logistic regression analysis by LeCroy et al. (1998). There are at least two potential explanations for this finding. On the one hand, this outcome was predictable from accepted sampling theory that underlies most advanced statistical tests like logistic regression. While the overall effects of specific risk factors on recidivism may be stable across study samples, the precise strength of the relationship represented by logistic regression estimates may nevertheless vary because of random sampling error. Based on this explanation, one would expect that precise probability estimates would not correspond to observed rates in most samples. On the other hand, this risk assessment study, like most, did not control for the effects of risk on interventions aimed to suppress recidivism rates for higher risk youths. Fewer than expected high risk youths may have re-offended because they received more intensive monitoring and intrusive interventions.

In response to these rather consistent findings across many studies, the trend in risk assessment design is to emphasize the classification function of risk assessment instruments rather than the predictive function. For example, rather than assigning a probability of recidivism to an individual or group of individuals, contemporary risk assessment instruments assign a score that reflects a greater propensity to re-offend than youths with a lower score. For the Arizona Risk/Needs Assessment Instrument, this would imply interpreting risk scores as an indicator of degrees of risk rather than as the probability of recidivism.

The third finding is in relation to the rate of incomplete needs scales (over 30% of adjudicated youths in FY04). The analysis suggests that needs scale completion rates are increasing but appear to fall short of the mandate suggested by policy. That needs scales are completed less often for low risk, non-severe offender groups, suggests the hypothesis that probation officers and juvenile court judges find information in the needs scale less relevant for these youths compared to higher risk, serious offenders. However, this hypothesis cannot be tested given available data. Unlike the risk scale, the needs scale does not include a measure to explain why the needs scale was incomplete for individual cases. In the present sample, missing needs assessment instruments for adjudicated youths precluded the planned analysis. Therefore, the utility

of the needs scale for enhancing the predictive validity of the Arizona Risk/Needs Assessment Instrument remains an area for further inquiry.

Finally, the fourth conclusion reflects on the potential of the risk scale to inform intervention planning. Overall, the study supports the use of risk assessment as a strategy to target higher risk offenders for more active interventions consistent with the "risk principle" (Bonta, 1996). In addition, the analysis of individual risk factors and recidivism suggests additional practice and policy implications. Among all risk factors, two stand out for the strength of their association with recidivism and for their relatively high prevalence in the sample: drug use and family relationship problems. This finding suggests the need to ensure that probation departments are adequately trained to address both problems and that probation departments have adequate resources to intervene into these problems when they occur. It should be noted that the presence of known family problems and drug use indicated by the risk assessment scale does not represent a full assessment of these issues. Further assessment is required to understand the mechanisms through which these two risk factors increase risk of recidivism for individual youths.

These findings suggest the following recommendations:

1. Changes to the risk scale are not recommended at this time beyond correcting the scoring error for 3+ time offenders. It should be noted that this correction preserves the integrity of the Arizona Risk/Needs Assessment Instrument but is not expected to alter the predictive validity of the instrument to any great degree.
2. Further investigation into the interpretation of risk scores by probation officers and judges is recommended to ensure that risk scores are not considered a deterministic probability for individual offenders but rather a relative indicator of youths' propensity to offend compared to youths with smaller risk scores. Subsequent to such an investigation, the department may wish to consider creating a three to five level classification system whereby youths would be grouped into risk groups ranging from "low risk" to "high risk." Such a change would ensure that risk scores were interpreted as relative to other youths rather than as deterministic of individual behavior.
3. Further investigation into the completion rates of needs scales for adjudicated youths is recommended to determine the extent of the problem, its probable causes, and potential remedies. Requiring that probation officers explain incomplete needs scales using a non-intrusive measure modeled after the risk scale would be an intermediate step toward understanding the problem. In addition, other potential barriers to completion should be explored including the availability of information needed to complete the needs scale, the length of the needs scale, and the utility of the needs scale for case planning.

References

Bonta, J. (1996). Risk-needs: Assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Catchpole, R. E. H., & Gretton, H. M. (2003). The predictive validity of risk assessment with violent young offenders: A 1-year examination of criminal outcome. *Criminal Justice and Behavior, 30*, 688-708.

Cottle, C. C., Lee, R. L., & Heilbrun, K. (2001). The prediction of criminal recidivism in juveniles: A meta-analysis. *Criminal Justice and Behavior, 34*, 367-394.

Dean, C. W., & Duggan, T. J. (1968). Problems in parole prediction: A historical analysis. *Social Problems, 15*, 450-459.

Funk, S. J. (1999). Risk assessment for juveniles on probation: A focus on gender. *Criminal Justice and Behavior, 26*(1), 44-68.

Gottfredson, D. M., & Tonry, M. (Eds.). (1987). *Prediction and classification: Criminal justice decision making*. Chicago: The Chicago University Press.

Gretton, H. M., McBride, M., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior, 28*(4), 427-449.

Griffin, P., & Bozynski, M. (2003). National overviews: State juvenile justice profiles.

Hoge, R. D. (2002). Standardized instruments for assessing risk and need in youthful offenders. *Criminal Justice and Behavior, 29*, 380-396.

Howell, J. C. (1995). *Guide for implementing the comprehensive strategy for serious, violent, and chronic juvenile offenders*. Washington, DC: Office of Juvenile Justice and Delinquency Prevention.

Howell, J. C. (2003). *Preventing & reducing juvenile delinquency: A comprehensive framework*. Thousand Oaks, CA: Sage.

Ilacqua, G. E., Coulson, G. E., Lombardo, D., & Nutbrown, V. (1999). Predictive validity of the Young Offender Level of Service Inventory for criminal recidivism of male and female young offenders. *Psychological Reports, 84*, 1214-1218.

Krysik, J., & LeCroy, C. W. (2002). The empirical validation of an instrument to predict risk of recidivism among juvenile offenders. *Research on Social Work Practice, 12*, 71-81.

LeCroy, C. W., Krysik, J., & Palumbo, D. (1998). *Empirical validation of the Arizona risk/needs instrument and assessment process*. Tucson, AZ: LeCroy & Milligan Associates, Inc.

Lipsey, M. W., & Derzon, J. H. . (1998). Predictors of violent or serious delinquency in adolescence and early adulthood. In R. L. D. P. Farrington (Ed.), *Serious & violent juvenile offenders* (pp. 86-105). Thousand Oaks, CA: Sage Publications.

Lipsey, M. W., & Wilson, D. B. (1998). Effective intervention for serious juvenile offenders. In R. Loeber & D. P. Farrington (Eds.), *Serious & violent juvenile offenders* (pp. 313-345). Thousand Oaks, CA: Sage Publications.

Rice, M. E., & Harris, G. T. (1995). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology, 63*, 737-748.

Schmidt, F., Hoge, R. D., & Gomez, L. (2005). Reliability and Validity Analyses of the Youth Level of Service/Case Management Inventory. *Criminal Justice and Behavior, 32*(329-344).

Schwalbe, C. S. (in-press). A meta analysis of juvenile justice risk assessment predictive validity. *Law and Human Behavior*.

Schwalbe, C. S., Fraser, M. W., & Day, S. H. (2007). Predictive validity of the Joint Risk Matrix with juvenile offenders: A focus on gender and race/ethnicity. *Criminal Justice and Behavior, 34*, 348-361.

Schwalbe, C. S., Fraser, M. W., Day, S. H., & Arnold, E. M. (2004). North Carolina Assessment of Risk (NCAR): Reliability and Predictive Validity with Juvenile Offenders. *Journal of Offender Rehabilitation, 40*, 1-22.

Schwalbe, C. S., Fraser, M. W., Day, S. H., & Cooley, V. (2006). Classifying juvenile offenders according to risk of recidivism: Predictive validity, race/ethnicity, and gender. *Criminal Justice and Behavior, 33*, 305-324.

Shlonsky, A., & Wagner, D. (2005). The next step: Integrating actuarial risk assessment and clinical judgment into an evidence-based practice framework in CPS case management. *Child and Youth Service Review, 27*, 409-427.

Silver, E., Smith, W. R., & Banks, S. (2000). Constructing actuarial devices for predicting recidivism: A comparison of methods. *Criminal Justice and Behavior, 27*, 733-764

Towberman, D. B. (1992). A national survey of juvenile risk assessment. *Family & Juvenile Court Journal, 43*, 61-67.

Appendix

A.1. Computation of predicted probabilities

A.2. Source table for Figure 1: 12-month new referral rates

A.3. Source table for Figure 2: Distribution of risk scores

A.4. Source table for figure 3: Distribution of recidivism by risk scores

A.5. Source table for Figure 4: Needs scale completion rates by risk level

A.1. Computation of predicted probabilities

IF Answered as noted		1st Referral		2nd Referral		3rd Referral	
Base Score			-1.4589		-.4621		.0263
Factors that affect score							
Current Offense is a Status Offense	**Y**	**Add**	**0.5160**	**Add**	**0**	**Add**	**0**
Juvenile's relationship with his/her family Involves frequent/intense conflict or is alienated/assaultive (known or suspected)	**Y**	**Add**	**0.6616**	**Add**	**0.5660**	**Add**	**0.3509**
Juvenile has ever been assaultive (answer or based on previous charges)	**Y**	**Add**	**0.5175**	**Add**	**0**	**Add**	**0**
Used or is suspected of using drugs Within the past year	**Y**	**Add**	**0.8060**	**Add**	**0.4701**	**Add**	**0.5619**
Ever truant or extensive absenteeism From school	**Y**	**Add**	**0.7392**	**Add**	**0.4835**	**Add**	**0.2328**
Currently enrolled in public, private or Home school regularly (This question was changed from analysis and is phrased in positive rather than negative "not enrolled")	**N**	**Add**	**0**	**Add**	**0.4350**	**Add**	**0**
Has behavioral problems/mental health issues	**Y**	**Add**	**0**	**Add**	**0.4337**	**Add**	**0**
Friends involved or suspected to be involved In delinquency	**Y**	**Add**	**0**	**Add**	**0**	**Add**	**0.3897**

Runaway, runaway attempts, known or Suspected	**Y**	**Add 0**	**Add 0**	**Add 0.3292**
Additional Prior Complaints*				**Add 0.4853**
SCORE FOR INSTRUMENT		**X**	**X**	**X**

Computation = $\frac{2.7182^{X}}{1 + (2.7182^{X})}$

A.2. Source table for Figure 1: 12-month new referral rates

	New referral rate (%)		
	FY02	FY03	FY04
Intake (%)	0	0	0
3-Month (%)	15	14	14
6-Month (%)	23	22	22
9-Month (%)	28	27	27
12-Month (%)	33	32	31

A.3. Source table for Figure 2: Distribution of risk scores

Risk Score	FY02	FY03	FY04
0-.20 (%)	36	38	39
.21-.40 (%)	15	15	16
.41-.60 (%)	17	17	18
.61-.80 (%)	19	17	16
.81-1.0 (%)	14	13	11

A.4. Source table for figure 3: Distribution of recidivism by risk scores

	Recidivism Rate (%)		
Risk Scores	FY02	FY03	FY04
0-.20	18	17	16
.21-.40	29	30	29
.41-.60	35	37	35
.61-.80	48	47	47
.81-1.0	55	53	52

A.5. Source table for Figure 4: Needs scale completion rates by risk level

	% Complete	
Risk Scores	FY03	FY04
.0-.20	46.9	40.3
.21-.40	77.8	78.9
.41-.60	75.4	77.8
.61-.80	66.2	75.1
.81-1.0	59.4	72.8