SUMMARY OVERVIEW
Submitted to The National Institute of Justice

Project Title: Evaluation of the Implementation of the Sex Offender Treatment Intervention and Progress Scale (SOTIPS)
Award #: 2012-AW-BX-0153
Final Submission Date: May 29, 2018

Authors:

University of Minnesota:

Michael H. Miner, Ph.D., Principal Investigator
Phone: 612-625-1500, Email: miner001@umn.edu

Beatrice "Bean" E. Robinson, Ph.D., Co-Investigator
Phone: 612-6248078, Email: robin009@umn.edu

Nicholas Newstrom, MA, Graduate Research Assistant
Phone: 612-625-1500, Email: newst038@umn.edu

Cathy Strobel-Ayres, BA, Research Coordinator
Phone: 612-624-7984, Email: cstrobel@umn.edu

Consultant:

R. Karl Hanson, Ph.D.
Phone: (613) 619-0817, Email: rkarlhanson@gmail.com

Grantee address:
University of Minnesota
Program in Human Sexuality
1300 2nd Avenue South, Suite 180
Minneapolis, MN 55454

This project was supported by Award No. 2012-AW-BX-0153, awarded by the National Institute of Justice, Office of Justice Programs, and U.S. Department of Justice. The opinions, findings, and conclusions, or recommendations expressed in this publication are those of the authors and do not necessarily reflect those of the Department of Justice.

Executive Summary

This 5-year study is set within the evolving field of sex offender risk assessment --- where research is moving beyond prediction based on static, unchanging risk factors to predictive tools that assess and identify risk and risk factors that serve as targets for therapeutic and supervision interventions. These dynamic risk factors form the basis of a number of recently developed tools, including the one tested here, the Sex Offender Treatment Intervention and Progress Scale (SOTIPS). This study has two parts: (1) a quantitative assessment of the psychometric properties, structure, and predictive validity of SOTIPS and (2) a formative evaluation of the implementation of SOTIPS in two geographically disparate major metropolitan areas --- Maricopa County (Phoenix and its surrounding suburbs) and New York City (NYC).

This study was designed to follow sex offenders on probation over time to assess recidivism and the predictive accuracy of SOTIPS. Actual data collection (date of first SOTIPS assessment thru date of last criminal record check) spanned 33.4 months (median = 34.3; *SD* = 8.8, range of 2-50 months). Probation officers in Maricopa County and NYC completed SOTIPS at three time points: at intake (the start of probation), 6 months after intake, and 12 months after intake, and the Static-99R once --- at the start of probation. The Static-99R is the most widely-used actuarial risk assessment tool for sex offenders. Data were collected on offenders' residence and employment stability, terminations or transfers of probation, attendance at sex offender specific treatment, and contacts between offenders and probation officers. Criminal record checks were run in the last year of the five year study --- mid-Year 5.

The results indicated that SOTIPS is a promising instrument for assessing dynamic risk factors in sex offenders on probation. It demonstrated good internal consistency and inter-rater reliability and significant incremental validity when combined with Static-99R. The addition of SOTIPS scores at baseline, 6-months and 12 months added information to the actuarial risk tool, Static-99R, and improved its predictive accuracy. The last SOTIPS, administered 12 months after the start of probation, was a better predictor of reoffending than the first SOTIPS, administered at the start of probation. This suggests that the best strategy would be to administer SOTIPS at regular intervals during supervision or treatment. If that is not possible, administering SOTIPS closer to release would result in better prediction.

The data were less clear about the factor structure of SOTIPS. There appeared to be some variance in the factor structure of SOTIPS over time and across the two study sites. We used four methods to select the number of factors included in the exploratory factor analyses, which resulted in slightly different results. Given that no single method has been shown to be consistently superior, looking for convergence amongst a number of methods is the recommended process. We did not find convergence across methods.

The formative evaluation of the implementation of SOTIPS provided important insights for future research and the process of implementing new procedures and processes within large complex governmental institutions. Participants in the initial focus groups identified common goals of collaboration and regular communication between probation officers and sex offender treatment providers. Participants in these initial focus groups also noted barriers within their administrative systems to such collaboration and communication. SOTIPS was implemented in New York City as planned, with the instrument completed by probation officers without input from treatment providers. Maricopa County experienced changes that interfered with the expected collaborations between probation officers and treatment providers. The major barrier was an increase in probation officer caseloads that stressed the system and limited the time and resources available to collaborate on SOTIPS assessments.

The perceptions of a sub-sample of offenders assessed in Maricopa County also provides insights into the probation and treatment system and the impact of implementing a risk assessment tool designed to track progress and identify intervention targets. Most offenders surveyed reported positive relationships with their probation officers and treatment providers. Results indicated that when an offender believed a particular problem area was important, he discussed this problem with his probation officer and treatment provider. As a group, the offenders studied were more likely to see themselves as low risk than the risk assessment tools indicated; risk assessment tools assigned substantial proportions of this sample to moderate or high risk groups.

There were several limitations to this study. First, the sample size in New York City was substantially smaller than expected. Less than 200 individuals were entered into the study, and this was likely due to smaller numbers of Static-99R eligible offenders being assigned to probation in lieu of prison. Because of this smaller than anticipated number, the recruitment time period in NYC was extended --- thus limiting the follow-up period for some of the NYC offenders. While criminal record checks were completed for most NYC offenders after 1–2 years at risk, a minority of NYC offenders were at risk for only a few months when the criminal record check was completed.

Data collection differed across the two sites. SOTIPS, Static-99Rs, and treatment progress reports were received from Maricopa County on paper forms and all other data were transferred as a data download from their probation activities data system. Probation officers in NYC completed study instruments on-line in a secure data management system. It is not clear whether these differences in data collection had any impact on the quality of the data.

Finally, the SOTIPS developers envisioned that SOTIPS would be completed by collaborating probation officers and treatment providers. In this study, that did not happen at either study site, although the initial intent in Maricopa County was to complete SOTIPS collaboratively. However, it does not appear that this process adversely affected the validity of SOTIPS. The

magnitudes of the single measure intraclass correlation and the average measure intraclass correlation were substantially the same --- indicating no noticeable advantage to more than one rater. Probation officer completed SOTIPS had good reliability, adequate predictive accuracy, and added significantly to the prediction of risk when used with the Static-99R.

In summary, this study revealed that SOTIPS is a promising dynamic risk assessment tool. Changes in SOTIPS scores improved prediction of risk when considered in conjunction with Static-99R; the total SOTIPS score formed a coherent scale; and two raters arrived at substantially the same score when scored within a month of each other. The finding that the most recent SOTIPS assessment was the most accurate indicates that probation officers can adjust their supervision levels based on current risk without compromising public safety. Instead of fixing supervision levels at the beginning of supervision, probation services can make more efficient use of their resources by routinely re-evaluating individuals and, if warranted, adjusting supervision levels. Basing decisions to change supervision levels on a validated dynamic risk tool, such as the SOTIPS, would increase the credibility of such risk management decisions.

Implementation of tools such as SOTIPS is a complex process that requires careful planning, buy-in from both the administrative structure and line staff, and adequate resources to insure that probation and treatment staff have the time and opportunity to collaborate. Jurisdictions that are thinking of using SOTIPS, or a similar tool, in conjunction with a static risk assessment tool must consider developing policies, guides and training for staff that help decipher results and commend changes based upon shifts in dynamic risk. Future studies should include a longer follow-up period that includes timeframes when individuals are no longer under probation supervision, larger samples to account for the low level of reoffending, especially sexual reoffending, and standard data collection procedures. Additionally, studies that employ implementation science could provide guidance on how best to adopt and use SOTIPS within complex probation and criminal justice systems.

Background

This study is set within the context of the developing sexual offender risk prediction field, where investigators are exploring reliable and valid means to assess what have been termed "dynamic risk factors." Instruments that identify the specific psychological risk factors present in the individual offender ought to allow treatment for that individual to be tailored to these specific needs, thus increasing treatment effectiveness. Thus, instruments have been designed to:

- Assess psychological factors that are empirically related to sexual recidivism, thus creating a basis for selecting treatment targets;
- Show incremental predictive validity relative to Static-99R or other measures of static risk factors;
- Measure change in a way that is related to sexual recidivism;

- Incorporate and point risk managers towards some of the factors identified in the desistance literature; and
- Improve the effectiveness of treatment in reducing sexual recidivism.

There have been at least three dynamic risk tools that have been developed: the STABLE/ACUTE (Hanson, Harris, Scott & Helmus, 2007), the Violence Risk Scale-Sex Offender Version (VRS-SO: Olver, Wong, Nicholaichuk, & Gordan, 2007), and the Sex Offender Treatment Intervention and Progress Scale (SOTIPS: McGrath, Lasher & Cumming, 2012). All of these tools and procedures have been developed to augment the findings from static measures and to account for and assess changeable factors (e.g., attitudes supportive of criminal behavior, impulsivity, interpersonal relationships, employment, etc.) that could be targeted by treatment and supervision interventions.

Preliminary research has suggested that SOTIPS is a promising instrument for use in treatment planning and risk management, and may be helpful in making decisions regarding treatment completion, reduced levels of supervision, and release from civil commitment and/or sex offender registration. McGrath et al. (2012) demonstrated that changes in SOTIPS scores were associated with reduced sexual recidivism and showed significant incremental predictive validity. McGrath et al.'s (2012) factor analysis of SOTIPS identified three factors: (1) *Sexual Deviance*, (2) *Criminality*, and (3) *Social Stability and Supports*.

The *Sexual Deviance* factor was composed of offense-related sexual interests, attitudes supporting sexual offending, and motivation to change sexual behavior. These elements were generally consistent with Mann, Hanson, and Thornton's (2010) meta-analysis of psychological factors predicting sexual recidivism. Ratings of stage of change, or readiness to make personal behavior changes (see Prochaska, Redding, Harlow, Rossi, & Velicer, 1994) were correlated with offense-supportive attitudes in factor analyses of the VRS-SO—suggesting that this element might be part of a broader attitudinal construct towards sexual offending.

The *Criminality* factor was composed of items expressive of general antisociality, impulsiveness, and oppositional reactions to rules; it was consistent with Mann et al.'s (2010) meta-analysis. The last factor, *Social Stability and Supports,* consisted of two constructs, dysfunctional coping (difficulties with problem-solving and emotion management) and the development of social support (e.g., residence, employment, and quality of social influences) (McGrath et al., 2012; Mann et al., 2010).

SOTIPS items appeared to be inclusive of the issues that have been found related to sexual offending and general criminality. Thus, SOTIPS is potentially useful for measuring dynamic factors that predict re-offending risk and can guide intervention, supervision, and dispositional decisions.

Since the McGrath et al. (2012) data was used to construct SOTIPS, its statistical properties may show shrinkage—a decrease in predictive and incremental validity. In order to demonstrate the predictive and incremental validity of SOTIPS, the findings of McGrath et al. (2012) need to be replicated. In addition, we need to better understand how SOTIPS can be used to guide treatment decisions, how to combine SOTIPS data with Static-99R data, and whether SOTIPS changes can be systematically used to change the level of risk predicted by a static, actuarial tool such as the Static-99R. This project was designed to study the predictive and incremental validity of SOTIPS in two sites that have not previously used the instrument and that are more ethnically diverse than the McGrath et al. (2012) sample.

Overall Project Design

The SOTIPS implementation evaluation began in spring 2013 with an initial planning meeting of the project investigators, study staff, and consultants to review objectives, outline data collection needs and methods, and discuss analysis and possible outcomes. Site meetings were held with key personnel in the probation departments at the two sites: Maricopa County in March 2013 and New York City (NYC) in April 2013. The focus of these meetings was to introduce the study team to each of the jurisdictions, to learn the characteristics and management systems at each site, and to outline the project goals and data collection requirements.

Next, SOTIPS group training sessions were conducted at each site by Dr. Robert McGrath, one of the developers of SOTIPS. In Maricopa County, both probation officers and treatment providers in contracted sex offender treatment programs were trained to score SOTIPS; in New York City, only probation officers were trained to score SOTIPS. Following the initial training, another "train the trainers" session was conducted at each site to ensure that as new staff was added, each site had sufficient numbers of trainers to train new staff to reliably score SOTIPS. Simultaneously, administrators from the nine sex offender treatment programs working with both probation sites were also contacted and asked to complete the North American Safer Society Foundation survey (McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010) describing their treatment models and programming (see attachment for treatment program profiles). In addition, two focus groups were conducted at each site—one with probation officers and one with treatment providers. All probation officers and therapists were invited to participate. Focus group questions concentrated on how information is shared and tracked, length and proportion of treatment, and how treatment progress is assessed. Focus groups were held again in 2017 to assess the effects of SOTIPS implementation on the exchange of information between probation officers and treatment providers and to identify any changes in the process of monitoring and making decisions about treatment completion.

Enrollment of sex offenders in the evaluation study began in April 2013. To be included, offenders needed to be Static-99R eligible (an adult male convicted of a contact or non-contact sex offense with an identifiable victim), mentally cognizant, at least 18 years old, and released to community supervision in January 2013 in Maricopa County and April 2013 in New York City.

Data collection differed at the two sites. In Maricopa County, data acquisition was delayed, due to delays in their implementation of a new data management information system (MIS). A new plan for data sharing between Maricopa County and the evaluation team was developed after Year 1. Maricopa Probation Department forwarded paper copies of the Static-99R, the three SOTIPS assessments, and treatment progress reports completed by the therapists to the evaluation staff who recorded these data. Maricopa County also provided data extractions from their internal probation tracking system (APETS) every 3 months. In New York City, data acquisition used REDCap, a secure data management system designed for multi-site studies. Headquarters staff entered demographic and background data when the offender was first assigned to adult probation. After their first meeting with the offender, probation officers took over data collection. Officers had unique log-in credentials and completed Static-99R, SOTIPS, 6-month progress reports, and reported any changes in probation status on-line. Static-99R and SOTIPS were automatically scored and officers could download and print copies of the completed instruments in .pdf format, as needed.

Probation officers completed the Static-99R and the first SOTIPS assessment at enrollment; a second SOTIPS assessment and a 6-month progress report six months after enrollment, and the third (and final) SOTIPS assessment and 6-month progress report one year after enrollment.[1] In addition, we collected basic offender demographic information (e.g., age, race, ethnicity, education, employment, etc.) and information regarding the index offense including their sex offense registration level, date of disposition (sentencing date), sentencing details, and victim demographic data. We collected Maricopa County demographic data via data extractions from the APETS system, while headquarters' staff in New York entered these data directly into the REDCap database

Sample Demographic Data

The overall sample had a mean age of 42.5 years (SD = 14.7). The New York sample was slightly younger (M = 40.8; SD = 16.2) than the Maricopa County sample (M = 43.0; SD = 14.1). Table 1 shows the race/ethnic breakdown of both samples, which were notably different. White, non-Hispanic offenders are the single largest race/ethnic group in the Maricopa County sample

[1] Six-month progress reports were only completed by New York City. These reports outlined current employment, relationship and housing status (including number of months at current address and number of family and/or others living with him), and number of probation officer visits (home, job, etc.). They also documented any new offenses, hiatuses (e.g., offenders out of supervision area or lost to supervision), and terminations from probation if any.

(54%), followed by Hispanics (27%). African American (36%) and Hispanic (34%) offenders were the largest ethnic groups in NYC; White, non-Hispanic offenders compose only 20% of the New York City sample. Another difference was that Native American offenders were more numerous in Maricopa County (4%) and Asian offenders were more prevalent in the New York City (5%) sample. Taken together, the sample was quite diverse. Additionally, the two sites also differed in terms of employment: most offenders in Maricopa County were employed and most offenders in New York City were unemployed. In both sites, the majority of the sample were single or dating, while less than a quarter were married or cohabitating.

Table 1. Demographic Information By Study Site

	New York City		Maricopa County	
	Frequency	%	Frequency	%
Race/Ethnicity				
Native American	1	.5	21	3.9
Asian	9	4.6	2	.4
African American	70	35.9	70	13
White	39	20	290	54
Hispanic	66	33.8	147	27.4
Other	10	5.1	7	1.3
Employment Status				
Employed	84	42.9	332	61.6
Unemployed	108	55.1	83	15.4
Retired			29	5.4
Disabled			39	7.3
Other/Unknown	4	2.0	56	10.3
Relationship Status				
Married/Cohabitating	45	23.0	129	23.9
Single/Dating	116	59.2	273	50.6
Other			111	20.6
Unknown	35	17.9	26	4.9

Results

This study was designed as a replication of McGrath et al. (2012) to explore the psychometric properties and predictive validity of SOTIPS. In addition, a formative evaluation of SOTIPS implementation was conducted, using focus groups, re-offense data, and probationer self-perceptions. In this section, we will present psychometric and predictive validity analyses and

discuss the implementation of SOTIPS within the two study sites as assessed by the qualitative data derived from the Year 1 and Year 5 focus groups; discuss changes in re-offending rates between a sample of offenders assigned to probation at the two study sites prior to implementation of this study and the rates found for the study sample; and discuss risk perceptions of a 10% sample of offenders drawn from Maricopa County.

Psychometric Properties and Structure of SOTIPS

The following sections report an analysis of the psychometric properties and structure of SOTIPS. The initial section will discuss the reliability of the instrument, both in terms of inter-rater reliability and internal consistency and the second section will address the structure of the instrument, using exploratory factor analysis.

Inter-rater Reliability

Each site was asked to double code a 10% sample of SOTIPS. This was done differently in the two sites. In Maricopa County, probation officers and contract treatment providers were trained to score and use SOTIPS. In New York City, only probation officers were trained to score SOTIPS. Therefore, inter-rater reliability will be reported separately for Maricopa County and NY.

Maricopa County.

A sample of 57 participants were selected for double coding by the individual's probation officer and treatment provider separately. The demographic characteristics of this sample mirrored those of the entire sample. There was some range in how close together the two ratings were, with a few as long as six months, but most within a month of each other.

New York City.

A sample of 20 offenders were selected for double coding. The supervising probation officer at each probation office coded selected offenders using the probation officer's case notes. Demographic characteristics of the sample mirrored those of the entire sample. Reliability coding was conducted an average of 3.9 months (SD = 2.2; median = 5 months) after the original coding. In fact, 40% of the re-coding was done 5 months after the initial probation officer SOTIPS coding. Coding lag time ranged from less than a month to 8 months.

Inter-rater reliability was calculated as the Intraclass Correlation Coefficient (ICC). Table 2 presents both the single measure and average measure ICC. The single measure reflects an estimate of the reliability of a single coder, while the average measure reflects the estimated reliability based on two coders and is commonly the ICC used for research purposes. The single measure is generally smaller than the average measure.

Table 2. Intraclass Correlations for SOTIPS Scores				
	Single Measure		Average Measure	
	ICC	95% CI	ICC	95% CI
New York City				
All Scores (N = 20)	.540*	.148, .787	.701*	.258, .881
Maricopa County				
All Scores (N = 57)	.653**	.475, .779	.790**	.644, .876
Scores within 2 months (n = 37)	.784**	.621, .882	.879**	.766, .937
Scores within 1 month (n = 26)	.821**	.644, .915	.902**	.783, .956

Note. ICC = Intraclass Correlation Coefficient. 95% CI = 95% Confidence Interval for the ICC.
* ICCs significant at p = .005
** ICCs significant at p < .001.

As can be seen in Table 2 the reliability estimates in New York were lower than those found in Maricopa County. However, they were not significantly different and the magnitude can probably be attributed to the differences in the recoding methods. In New York City, the SOTIPS recoding was done by Probation Office Supervisors after reviewing case files, while in Maricopa County, the reliability coding was done by two independent raters, the individuals probation officer and treatment provider, both of whom were familiar with the individual being assessed. Sample size was too small in New York City to look at ICCs with differing lag times, but in Maricopa County sample size allowed for exploration of how the time between coding affected the ICC. As would be expected, since SOTIPS should be sensitive to change over time, when ratings with longer time periods between them are excluded from analysis, the ICC increases, although there remains overlap in the 95% confidence intervals. Those SOTIPS scored within a month of each other show excellent inter-rater reliability. In all cases, there was no significant difference between the single measure ICC and the average measure ICC --- indicating that the score derived from a single coder provides reliable assessments.

Internal Consistency

Inter-rater reliability reflects whether, given the same information, two or more professionals will arrive at the same SOTIPS score. Another question is whether SOTIPS has internal consistency --- namely do the items that comprise SOTIPS form a coherent scale? Internal consistency was measured by coefficient alpha (α). Table 3 presents the estimated internal consistency reliability of the combined samples of SOTIPS at each of the three times that it was administered (intake, 6 months, and 12 months).

Table 3. Internal Consistency of SOTIPS at Each Administration				
	Sample Size	Coefficient Alpha (α)	Mean	*SD*
Time 1	696	.88	14.1[a]	8.4
Time 2	603	.90	12.5[b]	8.7
Time 3	540	.92	12.2[a,b]	9.2
a significantly different from b at p <. 05.				

The results in Table 3 indicate that SOTIPS has excellent internal consistency across all three administrations. In addition, mean SOTIPS scores at Time 1 were significantly higher than at Time 2. There was no difference between Times 2 and 3, and while Time 1 and 3 have about a 2-point difference, and Time 3 is essentially the same magnitude as Time 2, the difference between Times 1 and 3 do reach significance ($p = .067$). This is likely due to the slightly higher standard deviation in Time 3 scores.

In summary, SOTIPS showed good to excellent reliability whether measured by intraclass correlation or coefficient alpha. Thus, SOTIPS total score seems reliable and appears to maintain that reliability over time. In addition, the SOTIPS scoring protocol provided scores on five subscales that appeared to be meaningful. Previous research (McGrath, et al., 2012) found that SOTIPS was made up of three, not five factors. This 3-factor structure was found in the Vermont sample, which served as the developmental sample for SOTIPS; this project replicated the exploratory factor analysis and is discussed in the next section.

Structure of SOTIPS

Using exploratory factor analysis, McGrath et al. (2012) found that SOTIPS consisted of three factors: *Sexual Deviance, Criminality*, and *Social Stability and Supports*. In this study, exploratory factor analysis was conducted to replicate McGrath's analysis at three time points: initial/baseline assessment, 6-months, and 12-months after initial assessment. By conducting these analyses, the replicability of the McGrath et al. (2012) factor structure was explored and the stability of the SOTIPS factor structure over time was also investigated.

Exploratory factory analysis was conducted using MPlus8. Principle components analysis and Varimax orthogonal rotation were used to extract factors. Several methods have been suggested for determining the number of factors to retain in an exploratory factor analysis. The most common is Kaiser's criterion (Kaiser, 1960)—which involves keeping all factors with eigenvalues greater than 1.0. Parallel analysis is a newer method that has performed better in simulation studies and involves comparing the observed eigenvalues with reference eigenvalues from generated random data (Horn, 1960). Factors whose eigenvalues exceed the 95th percentile of their corresponding distribution of reference eigenvalues are retained—as long as they also exceed 1.0. A mathematically elegant process for determining the number of factors to retain in a principle components analysis is the Minimal Average Partial (MAP) test (Velicer, 1976). MAP calculates the average squared partial correlations for each factor after partialling out previously extracted factors. The number of factors retained corresponds to the number with the lowest average squared partial correlation. Simulations showed that the partial correlation initially decreases and then increases at some deflection point (Velicer, 1976). Each of these procedures have their strengths and weaknesses, and it is suggested that multiple procedures be used. Table 4 shows the numbers of factors that would be retained using each of the three methods over the three SOTIPS time periods.

Table 4. Results of Factor Selection Analyses			
	Kaiser	Parallel Analysis	Minimal Average Partial (MAP)
Time 1	4	3	2
Time 2	3	2	2
Time 3	2	2	2

Table 4 highlights two issues. First, each of the three methods suggested a different number of factors to be retained. Second, Time 1 SOTIPS showed the least consistent number of factors retained across the three time periods, MAP provided a consistent result across time, while Parallel analysis agrees with MAP at Times 2 and 3, but would provide a larger number of factors for Time 1 SOTIPS. Kaiser provides the least consistent result over time, with different numbers of factors retained at each of the three administration time points. Taken together, it appears that the most reasonable conclusion is to retain two factors. Table 5 shows the factor loadings for each of the factors for each administration generated by a principal components analysis with Varimax rotation.

Table 5. Factor Loadings across Time Frames						
	Time 1		Time 2		Time 3	
	Factor 1	Factor 2	Factor 1	Factor 2	Factor 1	Factor 2
Sexual Offense Responsibility	--	--	--	--	.385	--
Sexual Behavior		.420		.592		.598
Sexual Attitudes		.802		.866		.880
Sexual Interests		.839		.903		.905
Sexual Risk Management		.628		.722		.646
Criminal and Rule-Breaking Behavior		.602		.668		.742
Criminal and Rule-Breaking Attitudes	.926		.885			.876
Stage of Change	.937		.900			.871
Cooperation with Treatment	.391	.316	.556	.532	.632	.641
Cooperation with Community Supervision	.668		.719		.765	
Emotional Management	.753		.774		.695	
Problem Solving	.486	.517	.562	.548	.663	
Impulsivity	.626	.550	.659	.573	.706	
Employment	.600	.575	.591	.620		.658
Residence	.400		.554		.716	
Social Influences	.432		.444		.661	

Other than a minimal loading on Factor 1 at Time 3, it appears that *Sexual Offending Responsibility* had little relationship with any of the other items on the SOTIPS and did not load on either factor. *Sexual Behavior*, *Sexual Attitudes*, *Sexual Interests*, and *Sexual Risk Management* consistently loaded on Factor 2 across all time points, whereas *Cooperation with Community Supervision*, *Emotional Management*, *Residence*, and *Social Influences* consistently loaded on Factor 1 across all time points.

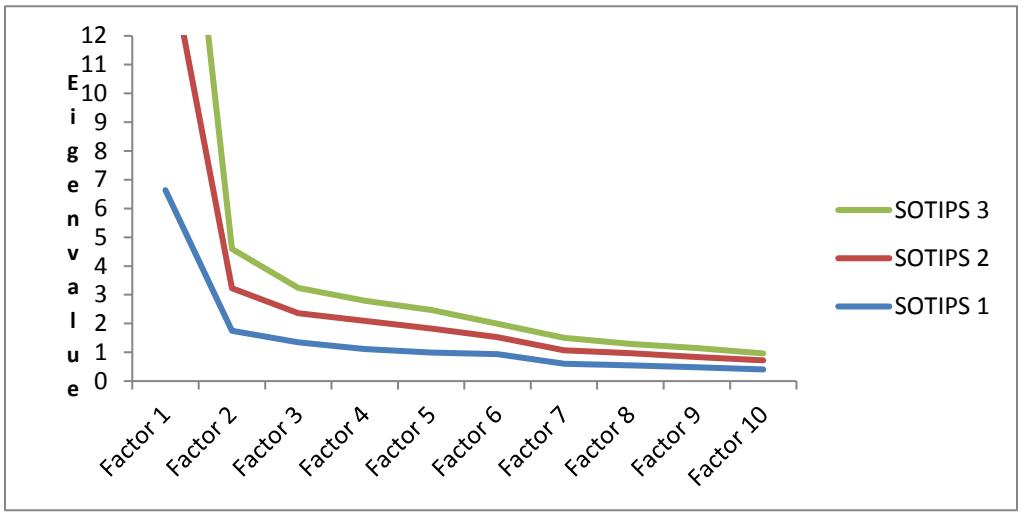
The other six items showed less consistent patterns. *Criminal and Rule-breaking Behavior* and *Stage of Change* showed high factor loadings on Factor 1 at Times 1 and 2, but equally high factor loadings on Factor 2 at Time 3. *Problem Solving, Impulsivity*, and *Employment* loaded substantially and equally on both factors at Time 1, but *Problem Solving* and *Impulsivity* loaded on Factor 1, whereas *Employment* loaded on Factor 2 at Time 3. *Cooperation with Treatment* loaded substantially and equally on both factors at all three time points. Recall that an orthogonal rotation was used—thus independent factors should have been extracted and there should be no item overlap. We see substantial item overlap, with 6 of the 16 items loading on both factors at least once and four items showing substantial overlap over at least two of the three time points. A recalculation of factor structure using an oblique rotation (Geomin) shows correlations between the two factors of 0.44 at Time 1, 0.60 at Time 2, and 0.66 at Time 3. In addition, while the fit statistics are not consistent, CFI and TLI indicate a good fit (greater than .90) across all three time periods (Table 6).

Table 6. Goodness of Fit Statistics for SOTIPS 2-Factor Solution

	RMSEA	CFI	TLI
Time1	.134	.928	.903
Time 2	.120	.955	.939
Time 3	.114	.971	.961

Tables 4-6 suggest that SOTIPS may be best conceptualized as two correlated factors. However, the inconsistency in the fit indices (Table 4) and the scree plot (Figure 1) called into question the two factor solution. Given these inconsistencies, we explored the results using a MAP analysis for each of the two study sites. We chose MAP since it provided the most consistent results over time and is the least subjective of the four criteria used in this study. In the Maricopa County data, MAP consistently indicated a single factor across all three time periods. In the New York City data, MAP indicated retention of three factors at Time 1, and two factors at Time 2 and 3. Thus, the MAP test indicated different principal component solutions across the two study sites, which further called into question the stability of the SOTIPS factor structure. At this point, there is inconclusive evidence for a multiple component structure for SOTIPS and given the high internal consistency of the scale, as measured by coefficient alpha, the predictive analyses used the total SOTIPS score. Additionally, our data suggested that *Sexual Offense Responsibility* could be removed from the instrument because it did not load on either factor when a two-factor solution was used and it significantly loaded only on a single factor at Time 3 (factor loading = .431).

Figure 1. Scree Plots for Each SOTIPS Administration



Predictive and Incremental Validity Results

SOTIPS shows good reliability, as measured by internal consistency and between raters. Thus, validity assessment is the next important step in determining how useful this instrument can be for probation supervision and sex offender treatment planning. In order for SOTIPS to provide useful information for decision-making, the predictive and incremental validity found by McGrath et al. (2012) requires replication. That is, SOTIPS needs to show the ability to predict re-offending and that it adds incrementally to static risk instruments in predicting re-offending at the same, or higher level than shown in the McGrath et al. (2012) developmental sample.

The recidivism prediction analyses were conducted on individuals with SOTIPS scores and recidivism data ($N = 717$). Participants had 1-3 SOTIPS assessments: 1st assessment ($n = 717$, $M = 14.1$, $SD = 8.4$); 2nd assessment ($n = 612$, $M = 12.5$, $SD = 8.6$); and 3rd assessment ($n = 543$, $M = 12.2$, $SD = 9.1$). Seven hundred and twelve participants also had a Static-99R score ($M = 2.3$, $SD = 2.1$). The average scores of both the SOTIPS and the Static-99R were similar to expectation for routine samples.

Actual data collection (date of first SOTIPS assessment thru date of last criminal record check) spanned 33.4 months (median = 34.3; $SD = 8.8$, range of 2-50 months). During this period, four types of recidivism were recorded: (a) conviction for a sexual offense (13/717 = 1.8%), defined as either contact (child molestation, sexual assault, attempted sexual assault, etc.) or a non-contact sexual offense (exposing, voyeurism, accessing child pornography, etc.) (b) charge or conviction for a sexual offense (16/717 = 2.2%); (c) any charge or conviction for a sexual or non-sexual violent offense (48/717 = 6.7%), defined as any sexual offense as defined above and violent offenses (assault, battery, robbery, etc.); and (d) any criminal recidivism (218/717 = 30.4%). Sexual offenses did not include Failure to Register. Our intent was to include only

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

criminal recidivism, not technical violations, in the final (any) recidivism category; however, criminal history data were augmented with data available from Maricopa County Adult Probations which often indicated an offender was in prison but did not indicate a charge. Thus, the nature of the charges was unknown in 160 of the 218 cases included in the any offense category. While it is likely that a significant proportion of the criminal recidivism events in this category might be related to violations of the conditions of probation, all of them were in the Maricopa County data and so were cross-referenced with the APETS data, which included information on probation violations. None of the 160 cases with unknown charges were reported as having violated probation in the APETS data. Each of the recidivism types were hierarchical, such that all sexual convictions were included in sexual charges, all sexual charges were included in sexual or violent offenses, and all sexual and violent offenses were included in the category of all criminal recidivism. Kaplan-Meier survival curves for each of the four outcomes are presented in Figures 2.

Figure 2. Kaplan-Meier Survival Curves for all Outcome Measures over 50 Months

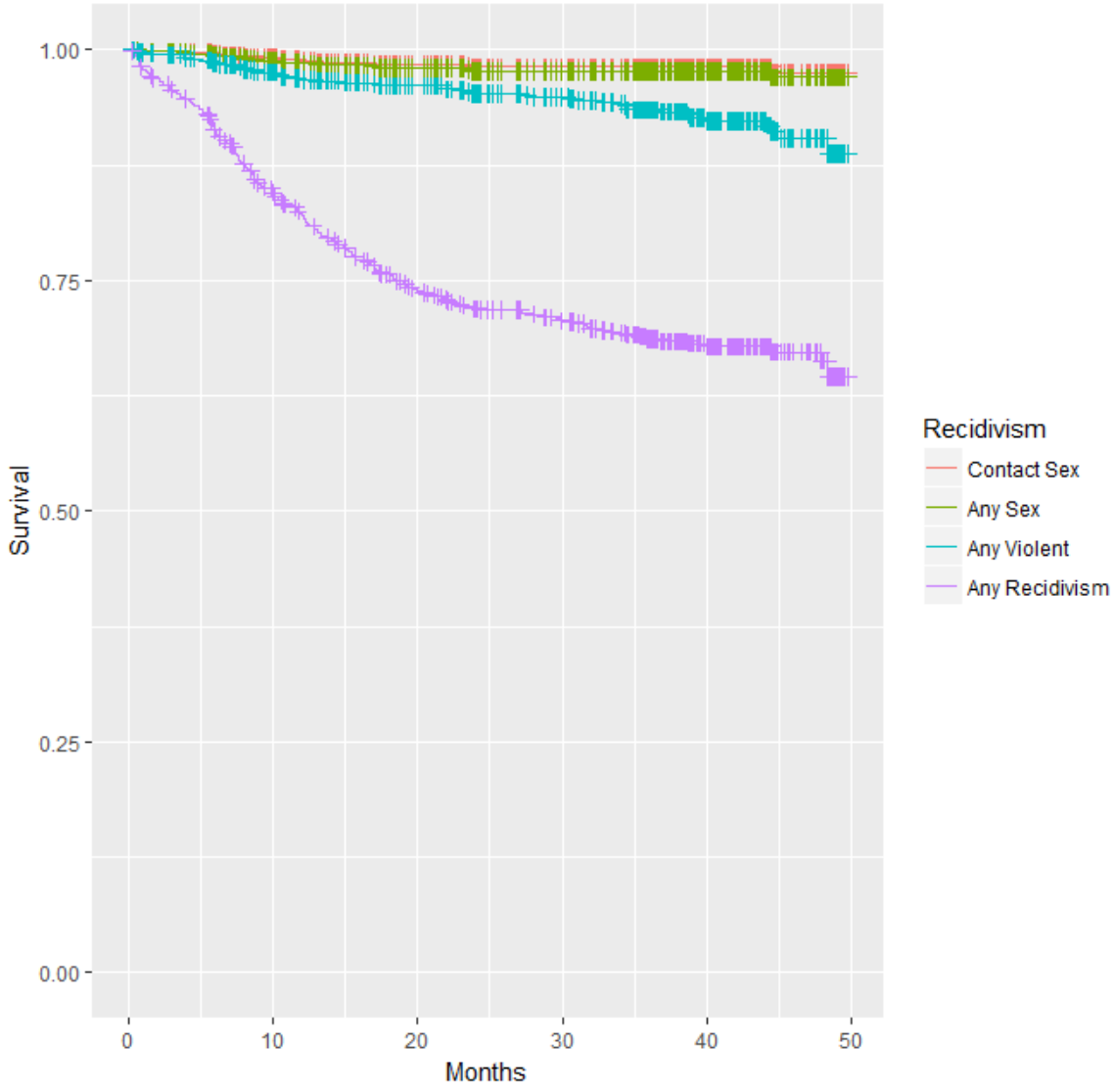


Table 7 presents the relationship of SOTIPS scores to recidivism at three time points. Based on AUC values, the relationship between SOTIPS was significant for 10 of the 12 comparison, with values ranging from .58 to .75. In comparison, the AUC values for Static-99R ranged from .59 to .69, and were statistically significant for three of the four comparisons. The AUC values are on the low end of the expected range for routine samples. The first SOTIPS assessment appeared to have less ability to predict sexual recidivism (AUCs of .63 and .64) than the third assessment (AUCs of .71 and .72). Please note the wide confidence intervals for most of the AUC analyses (with the exception of any criminal recidivism). With 5-16 sexual recidivists, the AUC values can change substantially based on a few cases (e.g., compare the AUC values for sexual convictions [n = 13] to the AUC values when three additional cases with charges are added [n = 16]).

Cox regression survival analysis (Singer & Willet, 2003) was used to compare the predictive accuracy (discrimination) of Static-99R, the first SOTIPS, and the three SOTIPS as a dynamic variable (time dependent covariate). The time dependent model used the last (most recent) SOTIPS score. Models were compared using the Akaike Information Criterion (AIC; Burnham & Anderson, 2004) for testing non-nested models. The AIC is computed based on the deviance (-2 log likelihood; -2LL) plus a penalty proportional to the number of parameters (*K*) used in the model. For the AIC, the penalty is twice the number of parameters (AIC = -2LL + 2*K*). Absolute AIC values are not interpretable. The difference between models, however, identifies the model that best fits the data, with low values indicating better fit. Although there are no absolute standards for evaluating differences, Burnham and Anderson (2004) interpret the difference between the minimum AIC and a model's AIC as indicating the degree of support for the model. They suggest that differences of less than 2 indicate substantial support (good agreement), differences of 4 – 7 as indicating a model has considerably less support than another, and models that are more than 10 AIC units higher than the minimum model as having "essentially no support." AIC is also used when comparing the difference between the fit of two models. AICs are interpreted similarly in this case, with differences of less than 2 indicating no difference between the model fits, 4-7 indicating modest differences and more than 10 indicating the model with the lower AIC is fits the data better than the other. Survival analyses were run using the Coxph program in R statistics (Therneau, 2015).

As can be seen in Table 8, the univariate effects were significant for both the first SOTIPS assessment and the dynamic SOTIPS for all outcomes. Although the dynamic SOTIPS was consistently more accurate than the first SOTIPS, the differences were small, and only meaningful for the outcome variable of *Any Criminal Recidivism*.

Static-99R showed predictive accuracy similar to that shown in other studies, with the exception of an unusually small (and non-significant) effect for sexual charges. In the Static-99R norms, the expected hazard ratio is 1.39 (95% CI of 1.33 to 1.45; Phenix, Helmus & Hanson, 2016) whereas the observed hazard ratio for sexual convictions in the current study was 1.35 (95% CI of 1.05 to 1.74).

For three of the four outcome variables, the model with the best fit combined Static-99R with the dynamic SOTIPS. This model had clear superiority for Any Criminal Recidivism (all AIC differences were greater than 10). Again, for sexual charges, adding Static-99R did not improve predictive accuracy over the dynamic SOTIPS (AIC difference of + 1.67).

Table 7. The relationship of Static-99R and SOTIPS assessments at three time points to subsequent recidivism.

Type of recidivism		SOTIPS						Static-99R	
		First		Second		Third			
		n	Mean (*SD*)	n	Mean (*SD*)	n	Mean (*SD*)	N	Mean (*SD*)
Sexual convictions	Yes	13	18.2 (9.1)	9	18.0 (5.0)	5	17.4 (6.4)	13	3.5 (1.4)
	No	704	14.1 (8.3)	600	12.4 (8.6)	539	12.1 (9.0)	699	2.3 (2.1)
AUC			.638 (.078)		.748 (.050)		.720 (.072)		.692 (.067)
95% C.I.			.485 -.791		**.650 - .847**		**.579 - .862**		**.560 - .824**
Sexual Charge	Yes	16	17.6 (8.3)	11	14.4 (5.3)	7	16.6 (5.6)	16	2.8 (2.1)
	No	701	14.1 (8.3)	598	12.4 (8.6)	532	12.1 (9.0)	696	2.3 (2.1)
AUC			.631 (.064)		.626 (.063)		.705 (.056)		.588 (.078)
95% C.I.			**.505 - .758**		**.503 - .748**		**.596 - .814**		.435 - .740
Violent or Sexual	Yes	48	16.4 (8.1)	33	14.2 (6.9)	24	13.4 (7.0)	48	3.0 (2.2)
	No	669	14.0 (8.4)	572	12.2 (8.6)	510	12.0 (9.0)	664	2.2 (2.0)
AUC			.597 (.041)		.601 (.044)		.579 (.052)		.614 (.044)
95% C.I.			**.517 - .678**		**.515 - .686**		.477 - .680		**.528 - .701**
Any criminal	Yes	217	17.8 (9.1)	114	16.6 (8.5)	59	15.3 (8.4)	215	2.9 (2.0)
	No	500	12.5 (7.5)	454	10.4 (7.3)	420	10.4 (8.1)	497	2.0 (2.0)
AUC			.677 (.022)		.717 (.027)		.683 (.036)		.633 (.023)
95% C.I.			**.634 - .721**		**.665 - .770**		**.611 - .754**		**.589 - .677**

Note. Confidence intervals that do not include .50 are in bold

Table 8. Cox regression analysis of Static-99R, first SOTIPS and dynamic SOTIPS for 712 individuals.

	B (SE)	e^B	e^B (95% CI)	C (SE)	-2LL, AIC	Δ AIC
Sexual Convictions (*n* = 13)						
Static-99R	.30 (.13)	1.35	**1.05 – 1.74**	.704 (.079)	162.77, 164.77	1.85
SOTIPS 1	.063 (.029)	1.065	**1.005 – 1.128**	.668 (.080)	163.68, 165.68	2.76
SOTIPS-D	.067 (.026)	1.070	**1.016 – 1.127**	.728 (.080)	162.79, 164.79	1.87
Static-99R	.24 (.14)	1.28	0.98 – 1.66	.760 (.080)	158.92, 162.92	lowest
SOTIPS	.056 (.027)	1.058	**1.003 – 1.116**			
Sexual Charges (*n* = 16)						
Static-99R	.13 (.12)	1.14	0.90 – 1.45	.599 (.071)	205.31, 207.31	4.39
SOTIPS 1	.056 (.027)	1.057	**1.003 – 1.115**	.662 (.072)	202.60, 204.60	1.68
SOTIPS-D	.060 (.024)	1.062	**1.013 – 1.114**	.717 (.072)	200.92, 202.92	lowest
Static-99R	.073 (.127)	1.08	0.84 – 1.38	.706 (.072)	200.59, 204.59	1.67
SOTIPS-D	.057 (.025)	1.058	**1.008 – 1.112**			
Violent or Sexual (*n* = 48)						
Static-99R	.20 (.071)	1.22	**1.07 – 1.41**	.626 (.043)	589.55, 591.55	3.89
SOTIPS 1	.048 (.016)	1.049	**1.016 – 1.083**	.629 (.043)	589.70, 591.70	4.04
SOTIPS-D	.045 (.014)	1.046	**1.017 – 1.076**	.673 (.043)	588.56, 590.56	2.90
Static-99R	.16 (.073)	1.18	**1.02 – 1.36**	.674 (.043)	583.66, 587.66	lowest
SOTIPS-D	.037 (.015)	1.038	**1.008 – 1.069**			
Any Crime (*n* = 216)						
Static-99R	.20 (.033)	1.22	**1.14 – 1.30**	.614 (.020)	2695.34, 2697.34	90.23
SOTIPS 1	.064 (.0072)	1.066	**1.051 – 1.082**	.663 (.020)	2660.02, 2662.02	54.91
SOTIPS-D	.073 (.0063)	1.075	**1.062 – 1.089**	.725 (.020)	2616.99, 2618.99	11.88
Static-99R	.13 (.035)	1.14	**1.06 – 1.22**	.732 (.020)	2603.11, 2607.11	lowest
SOTIPS-D	.067 (.0065)	1.069	**1.056 – 1.083**			

Note. Univariate effects are followed by the combined effect for Static-99R and dynamic SOTIPS (SOTIPS – D). Confidence intervals that do not include 1 are in bold.

Formative Evaluation of SOTIPS Implementation

The above sections provide support for the value of SOTIPS as a dynamic risk assessment tool. The data indicated that SOTIPS has adequate reliability and predictive validity. While there was mixed evidence for a multi-factor structure, the evidence indicated that the change in SOTIPS scores over time provided important additional information for assessing the risk posed by individuals on probation for sexual crimes. However, several additional questions are: how was SOTIPS implemented within the two study sites; were there important barriers to implementation; and did the probation and treatment systems change during the five-year duration of this project? These questions were addressed by three procedures: focus groups with probation officers and treatment providers in both study sites in Years 1 and 5, a study of a subsample of offenders drawn from Maricopa County and queried about their experiences in probation and treatment and on their perceptions of their reoffending risk, and a comparison of the reoffending rates of this sample with historical samples of men who meet inclusion criteria drawn from both sites.

Evaluation of SOTIPS Implementation and Effects on Probation and Treatment Systems

The evaluation of the SOTIPS implementation involved a qualitative analysis of focus group data collected at the beginning of the project, prior to SOTIPS training and implementation (Year 1) and then again in the last year of the project (Year 5). The purpose of the Year 1 focus groups was to assess the treatment and supervision system at baseline, while the Year 5 focus groups were designed to explore both probation officer and treatment therapist experience with SOTIPS, whether there were changes in how they collaborated and interacted, and whether there were changes on how decisions regarding treatment termination, changes in probation status, or other supervision decisions were made.

Maricopa County and New York City probation supervisors and program directors provided a list of all probation officers and therapists who supervised and treated probationers in their departments. All probation officers and therapists were contacted by email asking them to participate in a 2-hour focus group; the goal was to recruit 8-10 participants per focus group.

Two, 2-hour focus groups were held at each site in Years 1 and 5: one for probation officers and one for treatment providers (both groups included both frontline staff and supervisors). No monetary incentives were provided to participants, however, probation officer and treatment provider supervisors encouraged their attendance during their regular workday. Participants sat around tables and were digitally audio-recorded using Olympus DS-2 and DS-20 digital audio recorders with two external microphones. The moderators ended the focus groups on time after two hours. Research staff transcribed the Year 1 recordings from Maricopa County; the remaining Year 1 and 5 transcripts were prepared by Verbal Ink—a professional transcription service specializing in complex terminology and sensitive audio material. All focus group

participants gave first names to introduce themselves, enabling tracking of their comments throughout the focus groups. Identifiable information was removed upon transcription.

Year 1 Focus Groups: Table 9 displays the number of probation officers and therapists in each Year 1 focus group. The high numbers of participants made it difficult to get in depth information from each participant. In Year 5, recruitment was limited to avoid overly large groups and targeted 6-10 participants per focus group. We held Year 1 focus groups at the Maricopa County and New York City probation offices in April (Maricopa County) and August (NYC) of 2013.

Year 5 Focus Groups: Table 9 displays the number of probation officers and therapists in each Year 5 focus group. We held the Year 5 focus groups at the Maricopa County and New York City probation offices in January (Maricopa County) and May (NYC) of 2017. There were small numbers of therapists who participated in Year 5 focus groups; several therapists who signed up to attend were unable to attend for numerous reasons.

Table 9. Focus Group Numbers

Site	**N in Year 1 Focus Groups (2013)**		**N in Year 2 Focus Groups (2017)**	
	POs	**Therapists**	**POs**	**Therapists**
Maricopa County (incl Phoenix)	18	13	7	4
NYC	12	10	8	4
TOTAL	30	23	15	8

*Two of these participants were probation supervisors

Evaluators developed semi-structured parallel questions for probation officers and therapists in Year 1 and 5 in order to assess similar concepts. Questions included suggestions additional clarifying probes to encourage participants to elaborate on their responses (Appendices 2a and 2b).

Data Analysis

Qualitative analysis of the Year 1 focus groups was conducted by coding transcripts for themes in NVivo10. Coders used line-by-line open coding for all transcripts (Glaser, 1998). The coding strategy was based on social constructionism, a theory used to interpret reality out of the lived experiences of subjects.

Qualitative thematic analysis of the Year 5 focus groups was conducted by reviewing transcripts for answers to the specific questions noted in the grant proposal: how/if implementation of SOTIPS has changed the exchange of information between treatment providers and supervising agents; any changes in treatment process and acceptance by offenders; process of monitoring

offender progress through treatment; and decisions about treatment completion. Prominent themes with illustrative quotes were noted and described in results

Year 1 Focus Group Results

The initial focus groups discussed how information was exchanged between treatment providers and supervising agents, what vehicles existed for tracking offender progress in treatment, what measures were used to track offender progress, how the decision was made that an individual has completed treatment, what proportion of individuals who enter treatment actually complete treatment, and what is the average length of stay in treatment (see Appendix 2a).

Summary of Focus Group 1 Study Findings

Detailed findings from the Time 1 focus groups were presented in the following manuscript (Newstrom, Miner, Hoefer, Hanson, & Robinson, in press) and a summary of those findings germane to the formative evaluation are summarized here.

Positive collaborative relationships. Overall, both probation officers and therapists were quite positive about their collaborative partnership. The majority of probation officers and therapists agreed that regular, accurate, and timely communication between the probation officer and the treating therapist—preferably face-to-face—frequently occurred and was essential for providing good supervision and treatment to probationers. These findings confirm previous research on the importance of "frequent and substantive two-way communication between treatment providers and POs about information that will assist in reducing an offender's risk to the community" (McGrath, Cumming, & Holt, 2002, p. 62) and communication "as a pivotal concept and tool in developing positive inter-organizational relationships between community corrections and community treatment providers" (Monico et al., 2016, p. 497). Probation officers and therapists also emphasized how much they valued each other's roles—including appreciation for the fact that they believed that both probation officers and therapists usually operated within the established parameters of their professional responsibilities. The fact that probation officers and therapists were so positive about their relationships in spite of well-established differences between probation officer and therapist professional orientations cannot be taken for granted; probation officers and therapists have different professional perspectives that unless well-coordinated could present challenges to implementing collaborative approaches such as the containment model. When working together collaboratively, probation officers and therapists often need to set aside their preconceived ways of operating and integrate their different professional perspectives into their practices and relationships with their sex offender clients (D'Amora & Burns-Smith, 1999). Our findings suggest that the majority of the probation officers and therapists successfully navigated these differing professional perspectives and developed healthy collaborations.

Barriers to collaboration. On the other hand, not all probation officers and therapists who took part in the focus groups perceived their working relationships across disciplines as fully successful. These participants perceived barriers to collaboration due to infrequent, late, poor quality, and inaccurate communication; conflicts between the goals of therapy and probation; probation officers who did not value treatment or who were too tough, unrealistic, and confrontational; and therapists who were inexperienced, lacked sex offender specific training, did not value probation, or were not confrontational enough.

Some of the probation officers and therapists described troubled collaborative relationships partially due to differing professional perspectives. Such problems are not unique to the sexual offender field and reflect the common tension between the contrasting professional perspectives of law enforcement and treatment professionals in corrections and forensic mental health (Day, 2014; Hogue, 1993). Clear and Latessa (1993) described how conflict can exist between the probation officer's "law enforcer" role, which places emphasis on enforcing the legal requirements of supervision, and the therapist's "social worker" role, which places emphasis on assisting the offender with community adjustment and reintegration. Interestingly, Wilson and Draine (2006) observed greater collaboration between corrections and mental health when mental health services for offenders were led by corrections.

As the present results indicate, however, it is quite possible for these roles to be mutually enhancing. Many productive relationships, such as boss, coach, and parent, work best when they effectively combine support and control. The potential problems observed arose when the enforcement and support roles were poorly coordinated, or when one partner was not playing their expected role.

Additional Struggles Implementing Aspects of the Containment Model

Both sites reported that their supervision and management of sex offenders were based on the containment model (English, Heil, & Veeder, 2016). Probation officers and therapists discussed five deficits in resources and policies that interfered with adequate implement of the containment model: (1) high probation officer caseload size; (2) complexity and time intensiveness of GPS technology and monitoring; (3) inconsistent probation agency protocols/policies characterized by frequent changes in policy and the destabilizing effects of politics and politicians; (4) lack of a consistent funding source to pay for mandated sex offender treatment; and (5) inconsistent and absent actuarial risk assessment. Political influences were a major factor that affected these deficits.

Limitations

Krueger and Casey (2015) described the optimal size of non-commercial focus groups as 5-8 participants and 10-12 participants for marketing focus groups. The Maricopa County probation officer ($N = 18$) focus group was considerably larger than optimal, which may have made it more

difficult for Maricopa County participants to express their thoughts. To manage the higher numbers, we cut back on some of our questions and probes in order to give participants more time to answer questions and attempted to direct our questions to quieter probation officers. The space provided for the NYC focus groups had distracting environmental conditions (e.g., fan noise, and a lack of air ventilation / air conditioning) that may have interfered with participants' responses, including making it more difficult for participants and facilitators to hear each other, as well as affecting transcription quality.

As with any study involving a small sample of volunteers, it is unlikely that the participants were a random sample of any pre-defined population. Although we were fairly confident of the themes identified, it is possible that there were voices who did not find expression, either because they were systematically excluded (by themselves or others) or simply by chance. Consequently, it is unlikely that our results reflect the opinions and experiences of all therapists who treat sex offenders or probation officers who supervise such offenders in either location.

Another limitation is that some lines of questioning did not yield any interpretable findings. Specifically, respondents were unable to answer questions concerning the specific criteria they used to determine when offenders were ready to graduate treatment or to have reduced supervision requirements. This should not be entirely surprising because the scientific evidence on these topics is relatively weak and local policy is rarely specific about what constitutes an acceptable risk (see Hanson et al., 2017).

Summary and Conclusions

Overall, both probation officers and therapists were positive about their collaborative partnership in Year 1. The majority of probation officers and therapists agreed that regular, accurate, and timely communication between the probation officer and the treating therapist—preferably face-to-face—frequently occurred and was essential for providing good supervision and treatment to probationers. Probation officers and therapists also emphasized how much they valued each other's roles—including appreciation for the fact that they believed that both probation officers and therapists usually operated within the established parameters of their professional responsibilities.

On the other hand, not all probation officers and therapists perceived their working relationships across discipline as fully successful. These participants perceived barriers to collaboration due to infrequent, late, poor quality, and inaccurate communication; conflicts between the goals of therapy and probation; probation officers who did not value treatment or who were too tough, unrealistic, and confrontational; and therapists who were inexperienced, lacked sex offender specific training, did not value probation, or were not confrontational enough. Some of the

probation officers and therapists described troubled collaborative relationships partially due to differing professional perspectives.

In Year 1, probation officers and therapists expressed general dissatisfaction with the assessment instruments and process used to measure and track risk, progress, and recidivism potential, especially the variability of assessments, the imperfect science of assessments, and the need for the development of a more rigorous process for determining graduation criteria.

Year 5 Focus Group Results

The Year 5 focus groups concentrated on important issues/factors in supervising/treating sex offenders, indicators of supervision/treatment progress, how decisions are made around revocation, termination and release/graduation, working relationships with therapists (for probation officers) and probation officers (for therapists), important characteristics of high quality probation officer and therapist/ treatment provider relationships, and most importantly, the implementation, use, scoring, and efficacy of SOTIPS and how/whether the implementation of SOTIPS has changed the exchange of information between probation officers and therapists (Appendix 2b). Findings summarized in this report will focus on the implementation of SOTIPS and probation officer and therapist/ treatment provider relationships. We focus here on the issues of implementation and changes in relationships, in order to provide a process evaluation of the implementation of SOTIPS. We will do a more systematic coding of transcripts in order to determine whether an additional manuscript on the qualitative data is warranted.

Probation Officer and Therapist/ Treatment Provider Relationships in Year 5.
In Year 5, the majority of probation officers and therapists affirmed that factors identified in the initial focus groups as contributing to high quality collaborative relationships were still important. They stressed that regular, accurate, and timely communication between the probation officer and the treating therapist—preferably face-to-face—was essential for providing good supervision and treatment to probationers. Probation officers and therapists also emphasized the importance of both probation officers and therapists operating within the established parameters of their professional responsibilities.

However, there did appear to be increased tensions in the collaborative relationships between probation officers and therapists, especially in Maricopa County, for reasons that were not entirely clear. Probation officers and therapists pointed to several factors that might have contributed to these increased collaborative tensions:

- probation officers and therapists did not work together to score and use SOTIPS at either site,
- probation officer turnover,

- escalating probation officer caseloads in Maricopa County ("our numbers have overwhelmed us as a system"), and
- funding cuts to treatment programs in NYC ("a 75% decrease in funding").

Implementation, Use, Scoring, and Efficacy of SOTIPS

Most of the probation officers and therapists seemed to think they received good training in using SOTIPS—with the exception of several NYC probation officers. But there were numerous complaints about the implementation process. Several probation officers noted that SOTIPS was thrown at them, implemented without their input. They reported that—SOTIPS added yet another instrument to the stable of instruments they used already without thinking through which ones are really needed. This added more work on overloaded probation officer caseloads which one probation officer described as "paralysis by analysis." This contributed to feelings of resentment and the perception that SOTIPS is additional unnecessary paperwork. Both probation officers and therapists discussed how the initial excitement about the implementation of SOTIPS at all levels seemed to wane quickly over time.

Probation officers and therapists reported early attempts to work with each other collaboratively on scoring and using SOTIPS with little success. Probation officers discussed not getting return calls from therapists, giving up, and scoring the SOTIPS themselves. Therapists noted not having time to collaborate on the SOTIPS, disagreeing with probation officers over SOTIPS ratings, scoring it by themselves, and filing their assessment in the mental health records; most reported never being asked to share their SOTIPS ratings with probation. Most probation officers did not think that therapists used or scored the SOTIPS.

Notably, both probation officers and therapists noted the lack of administrative support, resources, and follow-up for the SOTIPS collaboration project (e.g., "it kinda fell by the wayside")—with many newer administrators not having much knowledge about the SOTIPS project. Several commented that these focus groups were the first time that they talked about SOTIPS since it was implemented five years ago.

On the other hand, the prevailing sentiment was that SOTIPS was easy to understand and was an effective and useful risk assessment tool because it was specifically geared toward sex offenders. Most felt that SOTIPS would be most useful if probation officers and therapists used the risk assessment collaboratively. However, they recognized the difficult logistics involved because of the size of caseloads and the difficulty in finding time to meet together, discuss and score it.

How/whether the implementation of SOTIPS has changed the exchange of information between probation officers and therapists

Probation officers and therapists reported that the implementation of SOTIPS fell short of its planned collaborative intent. Thus, the expectation or hope that the implementation of SOTIPS

would increase and improve the exchange of information between probation officers and therapist was not born out. In fact, disappointment with the lack of expected collaboration around SOTIPS may have contributed to the strained relationships between probation officers and therapists expressed during Year 5 focus groups. Participants reported that there were difficult logistics involved because of the increased size of caseloads in Maricopa County and the lack of administrative time and resources allotted to probation officers and therapists to collaborate on scoring and using SOTIPS together.

Conclusions

This research informs the supervision and therapy system conditions necessary for effective collaboration between probation officers and therapists. In particular, rules and procedures should seem reasonable, not arbitrary and rigid, and should fit into a coherent, coordinated vision. Effective collaboration between therapists and probation officers is possible, but it is fragile and should not be taken for granted. Those involved in these collaborations, and their administrators and systems, need to invest time and resources in building, monitoring, and maintaining these professional relationships. These collaborations work best when there are few barriers to informal communication, competent practitioners, respect for, and understanding of, each other's roles, and adequate time allotted and resources given to nurture these collaborative relationships. The results of the focus groups indicate that the administrative support and resources were not available to insure active collaboration between probation and treatment staff. In at least one site, barriers to such collaborations increased over the five years of this project and, thus, implementation of SOTIPS did not have the impact on collaboration that both administrative and line staff had expected.

Perceptions Project

The "*Perceptions Study*" was a part of the formative evaluation and was designed to assess treatment and probation interactions from the offender's perspective and their perception of their risk for reoffending. Data were collected by pencil and paper questionnaire from a sample of individuals recruited in Maricopa County, for whom we had Static-99R and SOTIPS scores. Data were collected in a group setting in four locations in the Phoenix area. Working with Maricopa County Adult Probation Department, 51 participants were recruited, about 10% of our sample. The *Perceptions Study* was not conducted in New York City since our total sample there was less than 200 leaving too few individuals from whom to recruit a sufficient number of participants for useful comparisons.

Participants in the *Perceptions Project* averaged 48.4 years of age (*SD* = 14.3), ranging from 24-73 years old. Sixty-nine percent of participants were White, 22% were Hispanic/Latino, and 9% were other race/ethnicity. One participant did not indicate race/ethnicity. Thirty-one percent of participants were married, 27% were single/never married, and 19% were divorced. Four

participants (8%) did not report relationship/marital status and the remaining 15% were separated, widowed, or cohabitating.

In addition to the individual's perceptions of their risk for reoffending, we collected additional information regarding their experiences with treatment and probation supervision. These data included the number of treatment sessions in the last month, the number of meetings with their probation officer in the last month, and the number of meetings where both were present in the last month. In addition, participants rated the importance of 15 topic areas (see Table 10) and whether they spoke with their probation officer or therapist about each. Participants also completed two scales, one assessing the quality of their relationship with their sex offender treatment provider and the other one assessing the quality of their relationship with their probation officer. These scales were adapted from the *STAR-P* (McGuire-Snieckus, McCabe, Catty, Hansson, & Priebe, 2007).

Table 10. Problem Areas	
1. Employment Status	9. Accepting responsibility for previous sex and non-sexual crimes
2. Substance abuse issues	10. High risk factors/Offending Cycle
3. Probation conditions (not including GPS)	11. Preparing for polygraph examination(s)
4. GPS Monitoring	12. Reviewing results of polygraph examination(s)
5. Family/Marital Relationship	13. Reviewing results of assessments (e.g. Static 99, ABEL)
6. Peer Relationships	14. Residence/Living arrangements
7. Leisure/Recreation Activities	15. Addressing mental health issues
8. Education Goals	

Results

The vast majority of participants (80%) met with their probation officer one time in the last month, with 6 participants meeting with their PO twice and 2 meeting with their PO weekly. Two-thirds (66.7%) of participants attended sex offender treatment weekly during the last month and 12% did not attend
treatment during the last month. Another 14% attended treatment between 1 and 3 times, while 6% attended 5 or more sessions during the last month. Only a small number of participants reported that they had a joint meeting with their PO and treatment provider (14%), with 10% having one meeting and another 4% having 5 or 6 meetings with both their PO and therapist.

Participants rated the importance of each of the problem areas shown in Table 10 on a 1 to 5 scale, where 1 indicated not important and 5 indicated very important. Most of the problem areas were rated as important, with only GPS monitoring and substance abuse issues having a

mean less than 3.0. There is little difference between the issues discussed with PO's and those discussed with therapists. Substantially the same proportions of participants indicated discussions with both PO's and therapists across all problem areas. Also, problem areas rated most important were also discussed with both PO's and therapists by larger percentages of the participants. Participants rated accepting responsibility for previous criminal behavior as most important.

Risk Perceptions and Instrument Assessed Levels

Participants were asked what risk level their PO and therapist had assigned them. Twelve percent of participants said that they didn't know, while 57% said they were assigned low risk, 26% moderate risk and 2 participants (5%) said they had been assigned as high risk. In terms of self-perception, a majority of offenders (96%) said they believed they were low risk for reoffending and 4% thought they were at moderate risk. No one believed they were high risk. The results from the Static-99R indicate that 60% of the participants scored in the low risk range, 30% in the moderate-low, 8 in the moderate-high range, and 2% (1 participant) scored in the high risk range. Participants Static-99R risk ranking was substantially the same as the Maricopa County sample, although 4% of the total study sample were categorized as high risk. Table 11 presents the risk categories as assigned by SOTIPS, as communicated to the participant by their probation officer and/or therapist and participants' self-perceptions. One participant did not have SOTIPS scores and 9 participants indicated that they did not know what risk level they had been assigned by their PO and/or therapist.

Table 11. Risk Levels Assigned by SOTIPS and Self-Perceived Risk Levels					
Risk Level	**SOTIPS 1**	**SOTIPS 2**	**SOTIPS 3**	**PO/TX Assigned***	**Self-Perception**
Low	24	28	26	29	49
Moderate	23	19	17	11	2
High	3	3	7	2	0

***Risk level communicated to participants by their PO and/or Therapist**

As can be seen in Table 11, risk levels as assigned by SOTIPS changed over time and, while similar to those assigned by POs and therapists, were quite different from participants' self-perceptions of risk. SOTIPS and PO/Therapist defined high risk was rare in participants, but substantial proportions of this sample had SOTIPS scores, and Static99-R scores in the moderate risk range. It appeared that, despite multiple sources of information on risk, many participants had unrealistic beliefs about their risk for re-offending. Given that only two participants rated their potential re-offense risk as greater than low, we cannot explore the impact that unrealistically low risk expectations have on treatment engagement, probation compliance, and reoffending.

One of the purposes of this project was to explore the effects of risk level labeling on probationers' behavior on probation and in treatment. In order to address this issue, we first ran an ANOVA exploring the effects that the level of risk assigned by probation officers and therapists had on the quality of the relationship with them as reported by participants. One would expect that if labeling was in operation, the assignment of a higher risk level would influence the quality of the relationship between offenders and authority figures. We found no significant difference across labeled risk level, although there were very few offenders in the high risk group ($n = 2$). We then combined the high and moderate risk groups and conducted a t-test. Again, there were no significant differences across the groups either on the quality of their relationship with their probation officer or the quality of the relationship with their therapist. There was no difference between the reported quality of relationships with probation officers and therapists, with medians of 29 and 30 respectively out of a maximum score of 44.

In order to better understand the effect that assessed risk, or treatment needs might have on the participants' reports of quality of relationship, we combined the high and moderate risk levels assessed by SOTIPS, since only 3 participants scored in the high risk category and conducted t-tests on the ratings of quality of relationship with probation officers and with therapists. The first SOTIPS administration showed no effect on quality of relationship with PO ($t_{48} = -1.55$, $p = .127$) or on quality of relationship with therapist ($t_{48} = -0.01$, $p = .993$). While not reaching conventional levels of statistical significance, there appears to be a trend for those rated as more at risk as rated by SOTIPS at time periods 2 and 3 ($t_{48} = -1.80$, $p = .078$ and $t_{48} = -1.93$, $p = .060$ respectively). The mean difference between groups at time period 2 was 2.6 and at time period 3

was 2.7, indicating that those categorized at lower risk experienced more positive relationships with their therapists. Relationships with probation officers did not differ at times 2 or 3.

Changes in Re-offending Rates

The data presented in this section serves two purposes: (1) to provide an estimate of the expected reoffending rate for the implementation sample, and (2) to serve as a baseline to determine whether any detected changes in the probation system over the 5 years of this project influenced rates of reoffending.

A sample of 500 men were selected in each site to match the inclusion criteria for the study. Thus, sampling criteria included being convicted of Static-99R eligible sexual offenses and placed on probation in either Maricopa County or New York City. The sample was drawn so that subjects would have at least 5 years' time at risk following being placed on probation. Thus, their convictions took place between 1978-2013.

Reoffending data were coded by two trained raters, with 23% double coded. Double coding was accomplished within a coding work group where a team came to a consensus as to how to code a random selection of criminal records. Data included arrest and convictions for contact sex offenses, non-contact sex offenses, non-sexual violent offenses, and non-sexual non-violent offenses. For each offense, the date of the offending behavior was coded, using date of arrest if date of offense was not available. The criminal history of each offender was completely reviewed to account for any incarceration time when calculating time at risk.

Data for Maricopa County were provided by the Maricopa County Adult Probation. The raw data consisted of Rap Sheets from a query of the Arizona State criminal history dataset for a random sample of 500 men. Data for New York City were provided by the New York State Bureau of Justice Statistics. The raw data consisted of a top charge query of the New York State criminal history dataset for a random sample of 501 men. New York City provided an SPSS dataset containing the criminal history data. Eleven offenders were removed from the sample because their crime was committed as a juvenile leaving a total sample of 490.

Only rudimentary demographics were available for this sample. Fifty-two percent were identified as Hispanic and 32% were Black. The only other substantial group was men identified as white, which represented 13% of the sample and a smaller percentage of Asian men (2%). The mean age of offenders at the time of their index offense was 35 years with a standard deviation of 17.4 years. Re-offenses occurred at a mean age of 36.5, which supports the observation that the majority of re-offenses take place shortly after the index offense.

Results

Data from both study sites were combined to determine the expected re-offending rate for the sample ($N = 968$) in each of the three offense categories. Over the 10 year follow-up period, 3.7% re-offended with a sexual offense, while 19% reoffended with a sexual or non-sexual violent offense, and 33% re-offended with any criminal offense. As noted earlier, the average follow-up period of the SOTIPS implementation sample was 33.4 months --- just under three years. To compare the re-offending rate of a sample of sex offenders drawn from our two study sites drawn prior to implementation of our study to that found through follow-up of the SOTIPS implementation sample, event history analysis was used to calculate re-offending rates at 2-year intervals. The 2- and 4-year re-offending rates for the baseline sample and the rates for the implementation sample are shown in Table 12.

Table 12. Re-offense rates for Baseline and SOTIPS Implementation Samples

Baseline	Sexual Crimes	Sex and Violent Crimes	Any Crime
0-2 Years	.02	.07	.09
2-4 Years	.01	.04	.06
Cumulative 4 year rate	.03	.11	.15
SOTIPS ($M = 33.4$ months	.02	.07	.30

The rates of criminal behavior between the baseline sample and the SOPTIPS implementation sample were comparable for sex and sex plus violent crimes, with SOTIPS implementation rates equal to the 2-year rates of the baseline sample and slightly lower than the 4-year rates. However, the SOTIPS implementation sample has twice the rate of any criminal arrest as the baseline sample.[2] The apparent increase in in any re-offending across our two study sites may indicate increases in such re-offending, or it may be due to differences in available data between the baseline data and the SOTIPS implementation sample data. Criminal history data were augmented by available probation department data in the SOTIPS implementation sample. This resulted in a substantial number of re-offenses in the SOTIPS implementation sample where the offense type was unknown. This was not the case with the baseline data since only criminal history data were used.

Discussion, Limitations, and Conclusions

During the last two decades, risk assessment has become integral to sexual offender management and correctional interventions. The most frequently used assessment tools include static unchangeable, factors that are mainly historical. Static factors, while useful for estimating long-term risk for behavior, are not helpful for developing targeted interventions. This project was designed to explore SOTIPS as a measure of dynamic risk for re-offending by men serving probation for sexual offenses.

[2] We had proposed to conduct a regression discontinuity analysis to determine how SOTIPS implementation affected re-offending rates. However, we do not have sufficient data points to conduct such an analysis, but the above data indicates little change in sexual and violent re-offending.

This project addressed two issues: (1) the structure, predictive validity and incremental validity of SOTIPS and (2) the effect that implementing SOTIPS had on probation and treatment systems in two geographically and demographically diverse jurisdictions - New York City and Maricopa County, Arizona - over time. While it was expected that SOTIPS would be scored collaboratively by probation officers and sex offender treatment providers, this was not what happened in practice. In both New York City and Maricopa County, SOTIPS was scored by probation officers with minimal, if any, input from treatment providers.

Structure and Validity of SOTIPS

Event history analysis (EHA) and receiver operating characteristic (ROC) both indicated good predictive validity for SOTIPS. Consistent with McGrath et al. (2012), this study's SOTIPS scores demonstrated AUC's consistently at .7 or above --- within the range found in a recent meta-analysis of numerous well known dynamic risk assessment tools (van den Berg et al., 2018). When combined with the Static-99R, SOTIPS scores demonstrated incremental validity. These results indicated that the initial SOTIPS added significantly to the prediction of all forms of re-offending when controlling for Static-99R score. Additionally, when treated as a time dependent variable, SOTIPS showed even better incremental validity beyond Static-99R, especially in the analysis of any criminal offense.

Analyses were inconsistent with respect to the factor structure of SOTIPS, however. Using principle components analysis, exploratory factor analyses were conducted on SOTIPS scores administered at intake, 6-months, and 12-months after intake. Varied factor identification criteria resulted in the identification of different numbers of factors, with MAP being most consistent in suggesting a two-factor solution. An orthogonal rotation failed to find a consistent second factor and certain goodness of fit indices indicated a two factor oblique solution might be a better fit for the data. However, different structures were obtained when controlling for study site, with MAP analyses consistently indicating a single factor across administrations in the Maricopa County data.

Evaluation of SOTIPS Implementation and Effects on Probation and Treatment Systems

Qualitative analyses indicated that while probation officers and treatment providers valued the roles of supervision and treatment respectively and saw collaboration as important for sex offender management, time constraints, caseloads, and differing priorities interfered with such collaborations. Not only did our focus group data indicate problems with collaboration, a 10% sample of probationers in Maricopa County indicated that it was rare for them to meet jointly with their probation officer and their treatment provider.

The formative evaluation, conducted using focus groups of probation officers and therapists in both study sites, indicated that SOTIPS implementation was problematic; line officers did not feel they were consulted about the use of the instrument and after initial implementation, there was little discussion of SOTIPS or its use. Both probation officers and treatment providers

expressed frustration with the lack of collaboration, although they also expressed the conviction that SOTIPS could be a useful tool, since it was specific to sexual offenders.

Offenders' Perceptions

We collected pencil and paper self-report data from a 10% ($N = 51$) sample of probationers in Maricopa County. Their perceptions of their risk differed from risk levels shared with them by their probation officers and treatment providers and from Static-99R and SOTIPS scores. The vast majority of offenders reported that they were at low risk for re-offending; only 2 of 51 offenders reported being at moderate risk and no one reported being at high risk. Scores on SOTIPS and Static-99R indicated more variability --- most scored as low risk, but some scored as moderate and a few scored as high risk. It appears that offender self-perception, at least on a self-report measure, differed from objective risk assessment measures and from the opinions of those providing supervision and treatment.

In general, the offenders who participated in this study reported fairly positive relationships with both their probation officers and their therapists. They tended to discuss those issues which they rated most important with both, and there were no differences in the issues discussed with probation officers and therapists.

Limitations

There were a number of issues that presented themselves during the course of this project that may limit the validity of results. While we proposed that data be collected from probation staff at both sites through an on-line data management system, this procedure was only implemented in New York City. Maricopa County intended to implement a new management information system (MIS) during the first year of this project and to provide the evaluation team with SOTIPS and Static-99R scores through data downloads from this system. However, during the second year of the project, it became clear that the MIS would not be implemented in time for use in this project. Maricopa County Probation Department staff collected and stored paper copies of SOTIPS and Static-99R forms and provided them to the evaluation team, who then checked for scoring errors and data entered completed instruments. When errors were encountered, evaluation staff sent forms back to Maricopa County for corrections. Ultimately, this delayed the timeline of the project in Maricopa County and caused some concern about the quality of the data.

The small sample size in NYC was another limitation. New York City was not able to enroll the 300-500 probationers called for by their implementation grant. Changes in New York State Criminal Statutes resulted in lower numbers of sexual offenders meeting Static-99R coding criteria and being assigned to community supervision during the time period covered in this study. In order to maximize sample size, enrollment in New York City was extended to the end of January 2015, instead of December 2013 as originally planned. Along with a relatively small

sample size of less than 200 participants, re-offending time at risk in New York City was also shortened; criminal record checks for recidivism were done for some offenders only a few months after completion of their 12-month SOTIPS. Issues with participant enrollment in New York City and data sharing in Maricopa County also required a change in the design of the criminal recidivism follow-up aspect of this study. The original design was to conduct two criminal record checks for recidivism: first at one year after 12-month SOTIPS and then again after three years. Due to time constraints, the first recidivism check was eliminated and only one was conducted. This criminal record check resulted in a time at risk of slightly less than 3 years in Maricopa County, but anywhere from a few months to three years after the 12-month SOTIPS was completed in New York City.

Analyses of predictive validity, especially when the *any criminal offense* category was used, may be limited by concerns regarding offenses with unknown charges. In order to better reflect the re-offending characteristics of this sample, we cross-checked data received from the search of criminal history databases in the two sites with data available from the sites' probation departments. This led to the identification of 160 cases where probation data indicated they were incarcerated, but did not indicate a charge, while the incident was not included in the criminal history data. Since we had no evidence that these cases were technical violations of probation conditions, which was also available in both the criminal history data and the probation department data, these cases were included in the any criminal offense category. This resulted in a more liberal definition of any criminal re-offending, thus a higher rate of recidivism.

There is likely some variation in the reliability of SOTIPS and Static-99R scores both within and across sites. Research has shown that while the Static-99R is less subject to individual bias, the Psychopathy Checklist-Revised (PCL-R), which requires more subjective application of criteria has been shown to be subject to bias in field studies (Murrie, Boccaccini, Guarnera, & Rufino, 2013). While there have been no studies of the extent to which instruments such as SOTIPS might be subject to bias, it is likely that the results would be more like those for the PCL-R than for Static-99R. Inter-rater reliability testing in New York City was less than optimal since there was no direct contractual arrangement between New York City Adult Probations and sex offender treatment providers. Thus, SOTIPS was only implemented in the probation department --- treatment providers did not systematically participate in SOTIPS training or scoring. Only the probation officer assigned to the probationer could complete SOTIPS. Supervisors did complete reliability SOTIPS on a subset of probationers, but they did this by relying on a review of case notes and records, which is a sub-optimal method for reliability coding (McGrath, private communication). Maricopa County provided better reliability estimates given that the probation officer assigned to an individual and their treatment provider completed SOTIPS independently. SOTIPS showed adequate to excellent inter-rater reliability in Maricopa County.

Conclusion

SOTIPS appears to be a promising instrument for measuring change in risk factors important to the prediction of sexual offending and other criminal behavior. Our analyses indicated that SOTIPS, when used with Static-99R, shows better predictive validity than either instrument alone. Further, SOTIPS appeared to be a dynamic instrument --- potentially useful in monitoring change that impacts risk. SOTIPS, therefore, may be a tool for decisions regarding supervision, treatment intensity and termination, and other decisions that are dependent on changes in risk level.

Implementation of an instrument such as SOTIPS must be done in a systematic manner, with input from the staff that are expected to score and use it. In this project, SOTIPS appears to have been viewed over time as another piece of paperwork, and did not have the impact on fostering collaboration and providing information to make decisions originally anticipated. SOTIPS did not replace tools already in use by the sites, but was an "add on" and neither probation officers nor treatment providers seemed to use the instrument as intended.

These data were insufficient to test the effect that implementing an instrument such as SOTIPS has on re-offending rates, treatment completion, or supervision failures. However, our focus group findings and the findings with respect to probationers' perceptions of risk would indicate little impact, at least as implemented for this project. SOTIPS has promise as a tool for assessing dynamic risk and changes in such risk over time. For this, or any tool, to have the desired impact, implementation needs to consider the workload, work flow, and decision-making processes of those charged with scoring, interpreting, and using the results. While SOTIPS implementation did not result in the desired collaboration, such collaboration appears to be important to and desired by probation and treatment personnel. Thus, implementation plans must consider the barriers to such collaboration.

In short, this project experienced a number of obstacles that affected the findings. However, the results indicated that even though the structure of SOTIPS is unclear, the instrument has predictive and incremental validity. Implementation was feasible within two very different settings and important barriers to effective implementation and collaboration were identified that can guide future research on this measure.

SOTIPS is a promising measure and its psychometric properties and predictive validity should be further replicated. Careful implementation within a system that facilitates the collaborative completion of the instrument by both probation staff and treatment staff would advance the understanding of how this instrument works as a facilitator of system interaction and as a predictor of risk and need. Further, replication of the factor analysis with a larger sample would clarify the structure of the instrument and might provide meaningful subscales that could be used

to better target interventions. Finally, predictive validation of SOTIPS requires implementation with a much larger sample and a much longer re-offense follow-up.

Jurisdictions that are thinking of using SOTIPS, or a similar tool, in conjunction with a static risk assessment tool must consider developing policies, guides and training for staff that help decipher results and commend changes based upon shifts in dynamic risk. The finding that the most recent SOTIPS assessment was the most accurate indicates that probation officers can adjust their supervision levels based on current risk without compromising public safety. Instead of fixing supervision levels at the beginning of supervision, probation services can make more efficient use of their resources by routinely re-evaluating individuals and, if warranted, adjusting supervision levels. Basing decisions to change supervision levels on a validated dynamic risk tool, such as the SOTIPS, would increase the credibility of such risk management decisions.

References

Burnham, K. P., & Anderson, D. R. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research, 33,* 261-304.

Clear, T. R., & Latessa, E. J. (1993). Probation officers' roles in intensive supervision: Surveillance versus treatment. *Justice Quarterly, 10*, 441-462.

D'Amora, D. A., & Burns-Smith, G. (1999). Partnering in response to sexual violence: How offender treatment and victim advocacy can work together in response to sexual violence. *Sexual Abuse: A Journal of Research and Treatment*, *11*, 293-304.

Day, A. (2014). Professional attitudes to sex offenders: Implications for multiagency and collaborative working. *Sexual Abuse in Australia and New Zealand, 6*(1), 12-19.

English, K., Heil, P., & Veeder, G. (2016). The Containment Approach: A strategy for the community management of sex offenders. In A. Phenix, & H. M. Hoberman (Eds), *Sexual offending: Predisposing antecedents, assessments and management* (pp. 713-744). New York: Springer Science.

Glaser, B. G. (1998). *Doing grounded theory: Issues and discussions*. Mill Valley, CA: Sociology Press.

Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*, *2007-05*. (ISBN No. 978-0-662-69992-7). Retrieved from https://www.publicsafety.gc.ca/cnt/rsrcs/pblctns/ssssng-rsk-sxl-ffndrs/ssssng-rsk-sxl-ffndrs-eng.pdf

Hanson, R. K., Bourgon, G., McGrath, R., Kroner, D., D'Amora, D. A., Thomas, S. S., & Tavarez, L. P. (2017). *A five-level risk and needs system: Maximizing assessment results in corrections through the development of a common language*. New York: The Council of State Governments Justice Center.

Hogue, T. E. (1993). Attitudes towards prisoners and sexual offenders. In N. C. Clark, & G. Stephenson (Eds.), *DCLP occasional papers: Sexual offenders* (pp. 27-32). Leicester: British Psychological Society.

Horn, J .L. (1960). A rationale and test for the number of factors in factor analysis. *Psychometrica, 30*, 179-185.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141-151.

Krueger, R. A., & Casey, M. A. (2015). *Focus groups: A practical guide for applied research, 5th Edition.* Thousand Oaks: SAGE Publications, Inc.

Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse*, *22*, 191-217.

McGrath, R. J., Cumming, G., & Holt, J. (2002). Collaboration among sex offender treatment providers and probation and parole officers: The beliefs and behaviors of treatment providers. *Sexual Abuse: A Journal of Research and Treatment, 14*, 49-65.

McGrath, R. J., Cumming, G. F., Burchard, B., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey.* Brandon, VT: Safer Society Press.

McGrath, R. J., Lasher, M. P., & Cumming, G.F. (2012). The Sex Offender Treatment Intervention and Progress Scale (SOTIPS): Psychometric properties and incremental predictive validity with Static-99R. *Sexual Abuse, 24*, 431-458.

McGuire-Snieckus, R., McCabe, R., Catty, J., Hansson, L., & Priebe, S. (2007). A new scale to assess the therapeutic relationship in community mental health care: STAR. *Psychological Medicine, 37*, 85-95.

Monico, L. B., Mitchell, S. G., Welsh, W., Link, N., Hamilton, L., Malvini Redden, S., ... & Friedmann, P. D. (2016). Developing effective interorganizational relationships between community corrections and community treatment providers. *Journal of Offender Rehabilitation*, *55*, 484-501.

Murrie, D. C., Boccaccini, M. T., Guarnera, L. A, & Rufino, K. A. (2013). Are forensic experts biased by the side that retained them? *Psychological Science, 24*, 1889-1897.

Newstrom, N. P., Miner, M. H., Hoefer, C., Hanson, R. K., & Robinson, B. B. E. (in press). Sex offender supervision: Communication, training, and mutual respect are necessary for effective collaboration between probation officers and therapists. *Sex Abuse.*

Olver, M. E., Wong, S. C., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale-Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment*, *19*, 318-329.

Phenix, A., Helmus, H., & Hanson, R. K. (2016). *Static-99R and Static-2002R evaluators' workbook*. Unpublished manual. Retrieved from www.static99.org

Prochaska, J. O., Redding, C. A., Harlow, L. L., Rossi, J. S., & Velicer, W. F. (1994). The Transtheoretical Model of Change and HIV prevention. *Health Education Quarterly, 21*, 471-486.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. Oxford, UK: Oxford University Press.

Therneau, T . (2015). A Package for Survival Analysis in S. Version 2.38 (software). Retrieved from http://CRAN.R-project.org/package=survival

van den Berg, J. W., Smid, W., Schepers, K, Wever, E. van Beek, D., Janssen, E., & Gijs, L. (2018). The predictive properties of dynamic sex offender risk assessment instruments: A meta-analysis. *Psychological Assessment*. *30*, 179-191.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321-327.

Wilson, A. B., & Draine, J. (2006). Collaborations between criminal justice and mental health systems for prisoner reentry. *Psychiatric Services*, *57*, 875-878.

Appendix 1

All of the sites provided services to adult males. Among the treatment sites, the range of offenders served varied, from 50 clients in New York to 1000 client served in Phoenix. All of the sites accepted self-payment; about half of the sites accept private insurance, a little over half of the practices (all of the Phoenix sites) use a form of government funding to treat offenders. However, only three sites provided treatment to adolescent offenders and only one Phoenix treatment provider served male children.

The average number of weekly treatment groups ranged from one group per week to one and a half groups per week. The length of these groups ranged from 60 to 120 minutes. Offenders met individually monthly for individual sessions compared to weekly for group sessions. One New York provider did not require offenders to meet at all individually while another provider reported that they see offenders individually on average, two times a month; the individual sessions ranged from 45-60 minutes.

Family and couples therapy were also provided to offenders. Four of the nine treatment providers did not offer these types of services to their clients. There was no consistent rate at which these services were offered; the length of these sessions ranged from 50-60 minutes.

Seven of the nine treatment providers conducted services at a private clinic. Other treatment settings include a community health setting and a court clinic. One treatment provider from New York and Phoenix reported that they have the ability to offer services at a probation office.

We asked treatment providers to estimate what percentage of their clients completed treatment. Answers ranged from 35% to 80 % success rate. Treatment providers also asked how long it takes clients to complete treatment. Estimates ranged from four to 40 months. Providers also were asked to report how long their step-down program is. Ranges reported were six months to 36 months. one Phoenix provider indicated that their supervision program was “unlimited” in length, suggesting that offenders can remain in the step-down program as long as needed.

The majority of treatment components provided by both sites focuses on interventions that address dynamic risk factors and provide insight into the offender’s behaviors. Addressing the client’s offense cycle, intimacy skills, problem solving, social skills and relapse prevention are universal treatments. Although not directly related to dynamic risk factors, helping the offender develop empathy for their victim(s) and responsibility for their crimes are treatment components that are universally included.

Table 1 Treatment components		
Therapy	***NY***	***PHX***
Assault cycle or offense chain	5	4
Trauma	5	4
Cognitive restructuring	5	3
Intimacy/Relationship skills	5	4
Offense responsibility	5	4
Problem solving training	5	4
Relapse prevention	5	4
Social skills training	5	4
Victim Awareness and empathy	5	4
Emotion Regulation	4	4
Sex education	4	4
Motivational interviewing	4	3
Victim clarification	4	3
Family reunification	3	4
Offense supportive attitudes	3	4
Self-monitoring training	3	3
Victim restitution	1	2
Schema therapy	1	1
Therapeutic community	1	0

We asked treatment sites to rank their top three treatment approaches to treating sex offenders. Approaches varied. However, cognitive behavioral therapy was the most commonly cited treatment approach across all sites. Other popular approaches included Risk, Need and Responsivity and Relapse Prevention models.

Appendix 2a
Year 1 Focus Group Questions

Probation Officer Questions	Therapist Questions
1. Do you supervise all sexual offenders in the same way, or do you do different things with different offenders? How do you decide what to do with different offenders?	1. Do you have one standard treatment program for each sex offender you treat or are there variations? On what basis do you make a determination as to what treatment program each sex offender client gets?
2. What do you focus on in sex offender supervision? What factors do you feel are most important to focus on in supervising sex offenders? What factors do you feel are of secondary importance in supervising sex offenders?	2. What do you focus on in sex offender treatment? What factors do you feel are most important to focus on in treating sex offenders? What factors do you feel are of secondary importance in treating sex offenders?
3. Do you have a method for deciding what to focus on in sex offender supervision for each client?	3. Do you have a method for deciding what to focus on in sex offender treatment for each client?
4. Do you create individualized or general supervision plans to reflect what the sex offender client needs to change?	4. Do you create individualized or general treatment plans to reflect what the sex offender client needs to change/treatment targets/identified needs?
5. How satisfied are you with the method you use to determine the focus of supervision/supervision plan for each sex offender client?	5. How satisfied are you with the methods you use to determine the focus of treatment/treatment plan for each sex offender client?
6. How do you track an offender's progress? How can you tell when and if they are getting better/ during the course of supervision?	6. How do you track a sex offender client's progress? How can you tell when and if they are getting better? Getting worse? What do you look for to determine their progress?
7. How do you generally feel about your method for tracking an offender's progress? What are its strengths and weaknesses? What would you change? Similar to what you said earlier?	7. How do you generally feel about your method for tracking a sex offender client's progress? What are its strengths and weaknesses? Similar to what you said earlier?
8. Most sex offenders have difficulties complying with all conditions of probation/parole. How do you determine when supervision is over—that it is time to revoke them or that they are doing well enough to relax the conditions of supervision? For example, when they violate their condition(s) by missing appointments with you, miss curfew, take a drink, become unreliable due to the chaos in their life—, how do you decide they've crossed the line or not?	8. How do you determine when sex offender treatment is over? For a successful completion? How can you tell/how do you decide that sex offender treatment has accomplished what it needs to? How do you decide that you've done enough?
9. How do you generally feel about your method for determining readiness for progress, revocation, or relaxing the conditions of probation? What are its strengths and weaknesses? Similar to what you said earlier?	9. How do you generally feel about your method for determining successful completion of probation from sex offender treatment? What are its strengths and weaknesses? Similar to what you said earlier?
10. I know this is hard to do, but could you make your best guess/estimate as to the percentage of offenders who complete supervision?	10. I know this is hard to do, but could you make your best guess/estimate as to percentage of offenders who successfully complete treatment?

11. What is your relationship with your client's treatment providers//therapists?	11. What is your relationship with your client's probation officers//supervising agents? How do you work with them? How involved are they in treatment?
12. How helpful is community supervision/your job in preventing sex offenders from reoffending?	12. How helpful is sex offender treatment/your job in preventing sex offenders from reoffending?
13. Have you heard or know about the "Sex Offender Treatment Intervention Progress Scale", commonly known as SOTIPS?	13. Have you heard or know about the "Sex Offender Treatment Intervention Progress Scale", commonly known as SOTIPS?

Appendix 2b
Year 5 Focus Group Questions

Probation Officer Questions	Therapist Questions
1. What factors do you feel are **most** important to focus on in supervising sex offenders? What do you focus on in sex offender supervision?	1. What factors do you feel are most important to focus on in treating sex offenders? What do you focus on in sex offender treatment?
2. What factors do you feel are **less** important to focus on when supervising offenders?	2. What factors do you feel are **less** important to focus on when treating offenders?
3. What has changed (if you feel things have changed) in how you do sex offender supervision in the **last 4 years** (since we were last here in 2013)? And what factors have led to these changes?	3. What has changed (if you feel things have changed) in how you do sex offender treatment in the last 4 years (since we were last here in 2013) and what factors have led to these changes?
4. Let's pretend that you go home tonight, you go to bed, and a miracle happens and the problems that one of your probationers has are miraculously solved. In this scenario, what are some changes you would expect to see in your probationer that give you confidence that they are unlikely to violate probation or reoffend?	4. Let's pretend that you go home tonight, you go to bed, a miracle happens overnight and the problems that one of your clients has are miraculously solved. What are some changes you would expect to see in your client that give you confidence that they have met all or most of their therapy goals and are ready to graduate from treatment?
5. How do you decide when to revoke an offender from supervision? For example, when they violate their condition(s) by missing appointments with you, miss curfew, take a drink, become unreliable due to the chaos in their life ---- how do you decide they've crossed the line or not?	5. How do you decide when a client has been unsuccessful in treatment and needs to be terminated from therapy?
6. Is there a process to release sexual offenders from probation when they are doing really well? If yes, how do you determine when an offender is doing well enough to relax their conditions of probation or release them from their supervision?	6. N/A
7. In what ways are treatment providers/ therapists involved in the supervision process? In what ways are you involved in the treatment process?	7. In what ways are probation officers involved in the therapy process? In what ways are you involved in the process of probation?
8. From the last focus groups we did here and in Phoenix/NYC in 2013, 4 important characteristics of high quality probation officer and therapist/ treatment provider relationships were identified; • regular, timely, accurate, face-to-face communication • probation officers valuing treatment and recognizing the importance of the treatment process • therapists valuing the role of probation in the management /treatment of sex offenders • both probation officers and therapists/ treatment providers knowing and adhering to their respective roles within the system.	8. From the last focus groups we did here and in Phoenix/NYC in 2013, 4 important characteristics of high quality probation officer and therapist/ treatment provider relationships were identified; • regular, timely, accurate, face-to-face communication • probation officers valuing treatment and recognizing the importance of the treatment process • therapists valuing the role of probation in the management /treatment of sex offenders • both probation officers and therapists/ treatment providers knowing and adhering to their respective roles within the system.

What do you believe are the most important characteristics of high quality relationships between probation officers and therapists/ treatment providers?	What do you believe are the most important characteristics of high quality relationships between probation officers and therapists/ treatment providers?
9. Who has used SOTIPS at least 1x since its implementation? *[show of hands]*	9. Who has used SOTIPS at least 1x since its implementation? *[show of hands]*
10. For those who used it, how was it scored? *[Prompt: Did you usually do it by yourself or with the client's therapist or did the probation officer complete it by themselves?]*	10. For those who used it, how was it scored? *[Prompt: Did you usually do it by yourself or with the client's probation officer or did the probation officer complete it by themselves?]*
11. Did you find it helpful scoring it by yourself or was it best scored with the probationers' therapist? *[Prompt: use response for question 10. Do they think it would have been more useful to score it with the therapist?]*	11. Did you find it more helpful to score it by yourself or was it best scored with the probation officer? *[Prompt: Do they think it would have been more useful to score it with the PO?]*
12. What can you tell us about how SOTIPS was implemented in the Maricopa County Probation Department? *[Probe: what was the process like? How well was it done?]*	12. What can you tell us about how SOTIPS was implemented in the New York City Probation Department? *[Probe: what was the process like? How well was it done?]*
13. SOTIPS asks questions about a variety of different areas of functioning. Which areas did you find most useful in your supervision of your probationers? ➢ 1. Sexual Offense Responsibility ➢ 2. Sexual Behavior ➢ 3. Sexual Attitudes ➢ 4. Sexual Interests ➢ 5. Sexual Risk Management ➢ 6. Criminal and Rule-Breaking Behavior ➢ 7. Criminal and Rule-Breaking Attitudes ➢ 8. Stage of Change ➢ 9. Cooperation with Treatment ➢ 10. Cooperation with Community Supervision ➢ 11. Emotion Management ➢ 12. Problem Solving ➢ 13. Impulsivity ➢ 14. Employment ➢ 15. Residence ➢ 16. Social Influences	13. SOTIPS asks questions about a variety of different areas of functioning. Which areas did you find most useful in your treatment of your clients? ➢ 1. Sexual Offense Responsibility ➢ 2. Sexual Behavior ➢ 3. Sexual Attitudes ➢ 4. Sexual Interests ➢ 5. Sexual Risk Management ➢ 6. Criminal and Rule-Breaking Behavior ➢ 7. Criminal and Rule-Breaking Attitudes ➢ 8. Stage of Change ➢ 9. Cooperation with Treatment ➢ 10. Cooperation with Community Supervision ➢ 11. Emotion Management ➢ 12. Problem Solving ➢ 13. Impulsivity ➢ 14. Employment ➢ 15. Residence ➢ 16. Social Influences
14. How did SOTIPS change (if it did change) the way information about probationers werewas shared between therapists and yourself?	14. How did SOTIPS change (if it did change) the way information about clients was shared between probation officers and yourself? 15. How did SOTIPS change (if it did change) the way information about probationers was shared between therapists and yourself?
15. In the focus groups we did 4 years ago, one of the issues that emerged was probation officer frustration with the constant changes in risk assessment tools that were not always helpful. How was SOTIPS more or less helpful in comparison to the other tools that you've been required to use? (e.g., Static 99R, LSIR, STABLE, etcetc.)	16. How was SOTIPS more or less helpful in comparison to the other risk assessment tools that you've used? (e.g., Static 99R, LSIR, STABLE, etc.).