———— **Research Report** ————

**Assessing Treatment Change
in Sexual Offenders**

Assessing Treatment Change in Sexual Offenders

Kevin L. Nunes

&

Franca Cortoni

Correctional Service of Canada

May, 2007

EXECUTIVE SUMMARY

The purpose of the current study was to assess the effectiveness of the National Sex Offender Program (NaSOP) by examining change on various treatment targets over the course of the program. These changes were assessed with dynamic risk assessment instruments and a number of self-report measures administered pre- and post-treatment. In addition to examining change, we assessed the validity of the measures by examining their correlations with validated risk assessment instruments. To the extent that the measures are correlated with risk, greater confidence can be had that the measures may be tapping criminogenic needs for sexual offenders.

Data were available for 313 sex offenders who had successfully completed the NaSOP. The majority participated in the moderate intensity program (75.4%). STATIC-99 scores were available for 225 participants. Based on STATIC-99 scores ($M = 2.88$, $SD = 2.03$), the average participant was moderate-low risk for sexual recidivism. SIR-R1 scores were available for 264 participants. Based on SIR-R1 scores ($M = 9.31$, $SD = 9.91$), the average participant was very low risk for general recidivism.

With regard to reliability, almost all of the self-report measures had acceptable internal consistency, indicating that the items of the measure were reasonably interrelated. Turning to concurrent validity, few of the self-report measures correlated with the risk assessment instruments; that is, the STATIC-99, the STABLE-2000, and the Violence Risk Scale: Sex Offender version (VRS:SO). Specifically, greater self-reported loneliness as measured by the UCLA Loneliness Scale was associated with greater risk as measured by the STABLE-2000. Greater physical aggression as measured by the Physical Aggression scale of the Aggression Questionnaire was associated with greater risk as measured by the STATIC-99 and the SIR-R1. Surprisingly, greater empathy for women in general was associated with *greater* risk as measured by the STABLE-2000, but the number of offenders in this analysis was very small ($n = 18$).

In terms of treatment change, small to medium sized improvements were observed on the dynamic risk assessment instruments and on almost all of the self-report measures over the course of the NaSOP. The improvements on the STABLE-2000 and the VRS:SO suggest that the NaSOP is successfully targeting dynamic risk factors for sexual recidivism.

The absence of an untreated comparison group does not permit strong conclusions about the causal influence of the NaSOP on the improvements observed. In addition, we were unable to examine whether treatment change was linked to reductions in recidivism because data on the NaSOP battery were unavailable for many offenders who had been released to the community.

Despite these limitations, the results of the current study are encouraging. As in research on other treatment programs for sexual offenders, improvement was observed on a number of areas targeted by the NaSOP. Positive change was observed on dynamic risk assessment instruments as well as on self-report measures. In combination with the significant reductions in recidivism associated with participation in the NaSOP (Cortoni & Nunes, 2007), the evidence suggests that the NaSOP is an effective program for sexual offenders.

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

The purpose of the current study was to assess the effectiveness of the National Sex Offender Program (NaSOP) by examining change on various treatment targets over the course of the program. These changes were assessed with dynamic risk assessment instruments and a number of self-report measures administered pre- and post-treatment. Although the most important outcome in evaluation of a sex offender treatment program is reduction of recidivism, it is also of interest to examine whether improvement has occurred on intermediate treatment targets. Focus on treatment change provides information about what aspects of the treatment program may be responsible for reductions in recidivism (Beech, Fisher, & Bishopp, 2004).

Recent meta-analyses suggest that sexual offender treatment programs are effective at reducing recidivism (Gallagher, Wilson, Hirschfield, Coggeshall, & MacKenzie, 1999; Lösel & Schmucker, 2005; Hanson et al., 2002; but see Rice and Harris, 2003 for a more cautious interpretation). Consistent with this research, we have found that sexual offenders who participated in the NaSOP had a lower rate of recidivism than those who did not participate in the NaSOP (Cortoni & Nunes, 2007). Statistically significant reductions in the rate of violent and any recidivism were observed. More specifically, the rate of violent recidivism was 83% lower and the rate of any recidivism was 77% lower in the NaSOP participants. Although reductions in the rate of sexual recidivism were also observed, this finding only tended toward statistical significance. Nevertheless, the rate of sexual recidivism was 68% lower in the NaSOP participants. This preliminary evidence suggests that the NaSOP is effective in reducing recidivism in sexual offenders.

The NaSOP is a cognitive-behavioural program offered at low and moderate intensity. The low intensity program is delivered both in the institutions and the community and primarily consists of 3 to 5 hours of group sessions per week over the course of 2 to 3 months, for a total of approximately 24 to 60 hours. The moderate intensity program is available in the institutions and primarily consists of 10 to 14 hours of group sessions per week over 4 to 5 months, for a total of approximately 160 to 280 hours. Both the low and moderate intensity programs are typically followed by maintenance sessions. The NaSOP is a therapeutic and semi-structured intervention aimed at reducing the risk of recidivism through the use of effective self-management. In addition to the general self-management objective, the program components are cognitive distortions, deviant arousal and fantasy, social skills, anger and emotion management, empathy, and victim awareness.

Some of these components of the NaSOP were assessed by the program test battery administered pre- and post-treatment. Change in risk for recidivism was measured with the STABLE-2000 (Hanson & Harris, 2000) and the Violence Risk Scale: Sex Offender version (VRS:SO; Wong, Olver, Nicholaichuk, & Gordon, 2000). Change was also assessed with self-report measures. Although improvement is often observed on self-report measures administered over the course of sex offender treatment programs (e.g., Beech, Fisher, & Beckett, 2004; Eastman, 2004), the available evidence is mixed in terms of the predictive validity of some of the dynamic factors targeted. In their recent meta-analysis, for example, Hanson and Morton-Bourgon (2004, 2005) provided evidence that sexual preoccupations, self-regulation problems, attitudes tolerant of sexual offending, intimacy deficits, and hostility were predictive of sexual recidivism. Other variables, however, such as lack of victim empathy and denial of sexual offences were weak predictors of sexual recidivism. In the current study, in addition to examining change, we assessed the validity of the measures by examining their correlations with validated risk assessment instruments. To the extent that the measures are correlated with risk, greater confidence can be had that the measures may be tapping criminogenic needs for sexual offenders.

METHOD

Participants

In the database were 439 male sex offenders who had participated in the low or moderate intensity sex offender program. Data on at least one pre- and post-treatment measure were available for 336 of these offenders. Of these 336, 313 (93.4%) successfully completed the program. Only these 313 offenders were included in the analyses below.

Measures

Risk Assessment Instruments

Static-99. The Static-99 (Hanson & Thornton, 1999) is an actuarial instrument designed to assess risk of sexual recidivism. It consists of 10 static items and scores can range from 0 to12. Higher scores reflect greater risk. Good predictive validity has been found for the Static-99 across several studies (Hanson & Morton-Bourgon, 2004).

Statistical Information on Recidivism – Revised (SIR-R1). The SIR-R1 (Standard Operating Practices [SOP] 700-04) is an actuarial instrument designed to estimate risk for general recidivism. The SIR-R1 consists of static items and *lower* scores are indicative of higher risk for recidivism. The SIR-R1 is currently used only with non-Aboriginal male offenders under federal jurisdiction. The

measure has demonstrated good reliability (internal consistency) and predictive validity for general, violent, and sexual recidivism in a variety of samples (Hanson & Morton-Bourgon, 2004; Nafekh & Motiuk, 2002).

Sex Offender Need Assessment Rating (SONAR) STABLE-2000. The STABLE-2000 (Hanson & Harris, 2000) is designed to assess dynamic risk factors for sexual recidivism. Six areas are assessed: Social Influences, Intimacy Deficits, Sexual Self-Regulation, Attitudes Supportive of Sexual Assault, Co-operation with Supervision, and General Self-Regulation. The total score can range from 0 to 12, with higher scores reflecting greater risk. Good predictive validity has been demonstrated in a prospective study (Hanson, 2005).

Violence Risk Scale: Sexual Offender Version (VRS:SO). The VRS:SO (Wong, Olver, Nicholaichuk, & Gordon, 2000) is designed to assess static and dynamic risk factors of sexual recidivism. In the version of the VRS:SO administered to participants, static risk factors were assessed with the Static-99, which was analyzed separately in the current report. The dynamic risk factors are Sexual Deviancy, Criminality, and Treatment Responsivity. Preliminary evidence supports the predictive validity of the VRS:SO (Olver, 2003).

Self-Report Measures

Balanced Inventory of Desirable Responding (BIDR). The Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1984) is a 40-item self-report measure designed to assess the tendency to respond to self-report scales in a socially desirable manner. The BIDR consists of two subscales: Impression Management (IM; 20 items) and Self-Deceptive Enhancement (SDE; 20 items). Scores on each subscale can range from 0 to 20. Higher IM scores indicate greater response bias due to a deliberate attempt to present oneself in a favourable light. Higher SDE scores reflect greater response bias due to self-deceptive overconfidence.

MOLEST and RAPE Scales. The MOLEST and RAPE scales (Bumby, 1996) were designed to assess cognitions supportive of, respectively, sexual abuse of children and sexual assault of women. On the MOLEST scale (38 items), scores can range from 38 to 152. On the RAPE scale (36 items), scores can range from 36 to 144. Higher scores reflect greater endorsement of cognitions supportive of sexual offending.

Miller Social Intimacy Scale (MSIS). The MSIS (Miller & Lefcourt, 1982) consists of 17-items. Scores can range from 17 to 170. Higher scores reflect higher levels of intimacy.

Revised UCLA Loneliness Scale. The Revised UCLA Loneliness Scale (Russell, Peplau, & Cutrona, 1980) consists of 20 items. Scores can range from 20 to 80 and higher scores reflect greater feelings of social dissatisfaction.

Child Molester Empathy Measure (CMEM). The CMEM (Fernandez, Marshall, Lightbody, & O'Sullivan, 1999) assesses empathy toward a) children in general, b) a victim of child sexual abuse, and c) one's own child victim(s). Scores can range from 0 to 500 on each subscale.

Rapist Empathy Measure (REM). The REM (Fernandez & Marshall, 2003) assesses empathy toward a) women in general, b) a victim of sexual assault, and c) one's own adult victim(s). Scores can range from 0 to 500 on each subscale.

Aggression Questionnaire (AQ). The AQ (Buss & Perry, 1992) consists of 4 scales. Scores can range from 9 to 45 on Physical Aggression (9 items), 5 to 25 on Verbal Aggression (5 items), 7 to 35 on Anger (7 items), and 8 to 40 on Hostility (8 items). Higher scores indicate greater aggression, anger, and hostility.

Sex Offender Acceptance of Responsibility Scales (SOARS). The SOARS (Peacock, 2000) was designed to assess the degree to which offenders accept responsibility for their sexual offences. Scores can range from 0 to 32 on Acceptance of Sexual Offense (8 items), 0 to 24 on Justifications for Sexual Offending (6 items), 0 to 32 on Acceptance of Offense Planning (8 items), 0 to 32 on Acceptance of Sexual Interests (8 items), 0 to 32 on Acceptance of Victim Harm (8 items), and 0 to 32 on Motivation to Change (8 items). Higher scores are indicative of greater acceptance of responsibility.

Procedure

The assessment battery was completed prior to participation in the program and, with the exception of the Static-99, again upon completion of the program. The data examined in this report were compiled from pre-existing databases, raw data, and file review. SIR-R1 data are not administered as part of the NaSOP battery; they were gathered from OMS for the current study.

RESULTS

Participants' mean age at the start of the program was 43.85 years (SD = 12.70). Mean sentence length was 3.84 years (SD = 2.50) for those serving a determinate sentence; 6.4% were serving an indeterminate sentence (e.g., life). Almost 8% of the participants were Aboriginal. Three quarters of offenders (75.4%) in this study participated in the moderate intensity program. It is noted that Static-99 scores were available for only 225 (out of 313) participants. Based on Static-99 scores (*M* = 2.88, *SD* = 2.03), the average participant was moderate-low risk for sexual recidivism. SIR-R1 scores were available for 264 participants. Based on SIR-R1 scores (*M* = 9.31, *SD* = 9.91), the average participant was very low risk for general recidivism.

Reliability

Internal consistency, one type of reliability, refers to the extent to which the items that make up a scale are related to, or consistent with, one another. To assess the internal consistency of the self-report measures in the NaSOP test battery, Cronbach's coefficient alphas were computed. Measures with missing items were excluded from all analyses. As shown in Table 1, most of the measures had acceptable internal consistency (Robinson, Shaver, & Wrightsman, 1991). The one exception was the Justifications scale of the SOARS, which had poor internal consistency at both pre- and post-treatment.

Table 1

Coefficient Alpha for Self-Report Measures

		Pre-treatment		Post-treatment	
Measure	Items	*N*	α	*N*	α
BIDR-SD	20	145	.69	136	.63
BIDR-IM	20	145	.79	136	.82
Bumby MOLEST	38	208	.95	230	.95
Bumby RAPE	36	229	.95	240	.96
Miller Social Intimacy Scale	17	270	.92	293	.93
UCLA Loneliness Scale	20	272	.89	285	.89
Child Molester Empathy Measure					
Children in general	50	165	.90	171	.92
Child victim of sexual abuse	50	165	.94	171	.92
Own child victim	50	165	.96	171	.95
Rapist Empathy Measure					
Women in general	50	93	.94	90	.95
Woman victim of sexual assault	50	93	.94	90	.91
Own woman victim	50	93	.96	90	.95
Aggression Questionnaire					
Physical	9	231	.79	225	.78
Verbal	5	231	.69	225	.67
Anger	7	231	.80	225	.76
Hostility	8	231	.81	225	.84
SOARS					
Acceptance of sexual offence	8	142	.80	135	.78
Justifications	6	142	.57	135	.54
Acceptance of offence planning	8	142	.84	135	.86
Acceptance of sexual interests	8	142	.79	135	.85
Acceptance of victim harm	8	142	.90	135	.79
Motivation to change	8	142	.75	135	.79

As shown in Tables 2 and 3, medium to large correlations with social desirability were found for almost all the measures. By convention, correlations[1] of around .10, .30, and .50 are respectively

---

[1] Correlations can be positive or negative depending on the direction of the relationship between the variables.

considered small, medium, and large effect sizes (Cohen, 1992). Despite these sizeable correlations, the BIDR was not included as a covariate in subsequent analyses for two reasons. First, because BIDR scores were unavailable for many participants, including it as a covariate would result in a loss of half of the participants in some cases and compromise statistical power. Second, there is evidence that socially desirable responding may itself be associated with risk. More specifically, greater social desirability has been associated with *lower* risk and *lower* likelihood of recidivism (Hanson & Wallace-Capretta, 2004; Mills & Kroner, 2005, 2006; Mills, Loza, & Kroner, 2003). Thus, correlations between a self-report measure and socially desirable responding may occur because both are related to risk and not because the validity of the measure has been compromised by social desirability. Based on these findings, Mills and colleagues (Mills & Kroner, 2005, 2006; Mills et al.) have argued that partialling out social desirable responding may yield misleading results with offenders. This argument is supported by the finding in the current study that the STABLE was significantly negatively correlated with social desirability at post-test (see Table 2).

Table 2

Intercorrelations between Risk Scales and Self-Deception (SD) and Impression Management (IM) of the Balanced Inventory of Desirable Responding (BIDR)

	Pre-treatment [a]			Post-treatment [b]		
	N	*r*		*N*	*r*	
Measure		SD	IM		SD	IM
Static-99	101	.08	.07	--	--	--
STABLE	62	-.00	-.13	62	-.35**	-.27*
VRS:SO						
Sexual Deviance	56	-.04	-.07	33	-.04	-.25
Criminality	56	.22	-.26[†]	36	.09	-.11
Responsivity	49	.18	.00	27	-.05	-.24

[a] Intercorrelations between pre-treatment BIDR and pre-treatment measures. [b] Intercorrelations between post-treatment BIDR and post-treatment measures.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3

Intercorrelations between Self-Report Measures and Self-Deception (SD) and Impression

Management (IM) of the Balanced Inventory of Desirable Responding (BIDR)

Measure	Pre-treatment [a]			Post-treatment [b]		
	N	r		N	r	
		SD	IM		SD	IM
Bumby MOLEST	114	-.25**	-.13	114	-.30**	-.03
Bumby RAPE	131	-.16[†]	-.08	123	-.25**	-.05
Miller Social Intimacy Scale	139	.33***	.23**	135	.18*	-.05
UCLA Loneliness Scale	137	-.44***	-.32***	128	-.46***	-.24**
Child Molester Empathy Measure						
Children in general	81	-.06	-.17	77	.05	-.01
Child victim of sexual abuse	81	-.08	-.23*	77	.10	.04
Own child victim	81	-.23*	-.18	77	.04	.01
Rapist Empathy Measure						
Women in general	34	-.10	.18	35	-.23	-.12
Woman victim of sexual assault	34	-.19	.22	35	.33[†]	.08
Own woman victim	34	-.29[†]	-.02	35	.05	-.16
Aggression Questionnaire						
Physical	124	-.14	-.25**	99	-.20*	-.35***
Verbal	124	-.26**	-.37***	99	-.10	-.28**
Anger	124	-.33***	-.35***	99	-.30**	-.32**
Hostility	124	-.48***	-.43***	99	-.39***	-.35***
SOARS						
Acceptance of sexual offence	118	-.34***	-.24**	100	.04	-.12
Justifications	118	-.21*	-.14	100	-.03	.12
Acceptance of offence planning	118	-.30**	-.19*	100	-.11	-.05
Acceptance of sexual interests	118	-.20*	-.30***	100	-.03	-.24*
Acceptance of victim harm	118	-.18*	-.16[†]	100	.03	-.19[†]
Motivation to change	118	-.31***	-.12	100	-.07	-.17[†]

[a] Intercorrelations between pre-treatment BIDR and pre-treatment measures. [b] Intercorrelations

between post-treatment BIDR and post-treatment measures.

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

As shown in Table 4, the intercorrelations between the risk assessment instruments were

generally as would be expected. Few of the self-report measures (Table 5), however, correlated with

the risk assessment instruments. Greater self-reported loneliness as measured by the UCLA Loneliness Scale was associated with greater risk as measured by the STABLE-2000. Greater physical aggression as measured by the Physical Aggression scale of the Aggression Questionnaire was associated with greater risk as measured by the Static-99 and the SIR-R1. The findings for the Rapist Empathy Measure were somewhat surprising. Greater empathy for women in general was associated with *greater* risk as measured by the STABLE-2000. The results for empathy for one's own victim were more consistent with expectations. Specifically, less empathy for one's own victim was associated with greater risk on the STABLE-2000; this correlation, however, failed to reach statistical significance. Due to the small number of participants in these correlations between the Rapist Empathy Measure and the STABLE-2000 ($N = 18$), these results should be interpreted with caution.

Table 4

Concurrent Validity of the Risk Scales and Change Measures at Pre-Treatment

Measure	Static-99		STABLE-2000		SIR-R1	
	N	*r*	*N*	*r*	*N*	*r*
Static-99	--	--	--	--	193	-.55***
STABLE	101	.19†	--	--	89	-.01
VRS:SO						
Sexual Deviance	132	-.05	80	.41*	112	.20*
Criminality	131	.31*	78	.33*	111	-.55*
Responsivity	117	.05	71	.20†	100	.05

$^\dagger p < .10.$ * $p < .05.$ ** $p < .01.$ *** $p < .001.$

Table 5

Concurrent Validity of the Self-Report Scales and Change Measures at Pre-Treatment

Measure	Static-99		STABLE-2000		SIR-R1	
	N	*r*	*N*	*r*	*N*	*r*
Bumby MOLEST	145	-.06	80	.14	173	.13†
Bumby RAPE	158	-.07	76	.06	197	.08
Miller Social Intimacy Scale	188	-.07	91	-.08	228	.04
UCLA Loneliness Scale	193	-.03	93	.24*	227	.02
Child Molester Empathy Measure						
Children in general	113	-.18†	66	.08	132	.01
Child victim of sexual abuse	113	-.09	66	-.01	132	.04
Own child victim	113	-.06	66	-.11	132	.13
Rapist Empathy Measure						
Women in general	68	-.18	18	.51*	77	.00
Woman victim of sexual assault	68	-.05	18	.17	77	.00
Own woman victim	68	.04	18	-.24	77	.11
Aggression Questionnaire						
Physical aggression	166	.27*	76	-.01	196	-.32*
Verbal aggression	166	-.04	76	.05	196	.04
Anger	166	.09	76	.16	196	-.09
Hostility	166	-.08	76	.19†	196	.07
SOARS						
Acceptance of sexual offence	89	-.04	62	.15	125	.10
Justifications	89	.12	62	.09	125	-.13
Acceptance of offence planning	89	.20†	62	.13	125	-.03
Acceptance of sexual interests	89	.16	62	.09	125	-.03
Acceptance of victim harm	89	-.05	62	-.06	125	.06
Motivation to change	89	.06	62	.13	125	.13

†*p* < .10. * *p* < .05. ** *p* < .01. *** *p* < .001.

As shown in Tables 6 and 7, the within-subjects analyses of variance (ANOVAs) comparing the pre- and post-treatment scores indicated statistically significant improvement on almost all the measures. Because within-subjects analyses are very high in statistical power, however, inspection of the effect sizes (correlation coefficients) in Tables 6 and 7 is informative. The effect sizes indicate the magnitude of the change from pre- to post-treatment. Although the comparisons are within-

subjects (i.e., comparing the same participants on a measure administered at different times), the correlations were computed as if they were between-subjects (i.e., comparing different participants) to avoid inflating the estimate of the magnitude of change. Thus, correlations computed as though they were from separate participants yielded correlations that more accurately reflect the magnitude of change from pre- to post-treatment.

As shown in Tables 6 and 7, small improvements (ranging from .09 to .20) were observed on the Criminality factor of the VRS:SO; UCLA Loneliness Scale; empathy for children in general on the CMEM; empathy for women in general on the REM; Physical Aggression and Hostility scales of the Aggression Questionnaire; and Justifications, Acceptance of Sexual Interests, and Motivation to Change scales of the Sex Offender Acceptance of Responsibility Scales.

Medium improvements were found (ranging from .24 to .39; see Tables 6 and 7) on the STABLE-2000, the Sexual Deviance and Responsivity factors of the VRS:SO; Bumby MOLEST and RAPE scales; empathy for a child victim of sexual abuse and empathy for one's own victim of sexual abuse on the CMEM; empathy for a woman victim of sexual assault and empathy for one's own victim of sexual assault on the REM; and the Acceptance of Sexual Offence, Acceptance of Offence Planning, and Acceptance of Victim Harm scales of the SOARS.

Table 6

Change on Dynamic Risk Scales

Measure	N	Pre-Program		Post-Program		F	r
		M	SD	M	SD		
Static 99	225	2.88	2.03	--	--	--	--
SIR-R1	264	9.31	9.91	--	--	--	--
STABLE-2000	87	5.57	2.13	4.49	2.28	24.10***	.24
VRS:SO							
Sexual Deviance	80	6.99	3.63	4.82	3.70	99.90***	.28
Criminality	82	7.11	4.33	5.47	4.54	61.43***	.18
Responsivity	72	5.36	2.01	4.20	2.59	55.49***	.24

$^{\dagger} p < .10. * p < .05. ** p < .01. *** p < .001.$

Table 7

Change on Self-Report Measures

Measure	N	Pre-Program		Post-Program		F	r
		M	SD	M	SD		
BIDR							
Self-deception	120	5.58	3.20	5.68	3.02	0.14	-.02
Impression management	120	6.27	4.00	6.73	4.45	2.50	-.05
Bumby MOLEST	198	64.81	17.53	53.87	14.83	125.20***	.32
Bumby RAPE	214	60.14	15.23	49.22	13.99	174.39***	.35
Miller Social Intimacy	262	135.14	23.73	135.98	25.25	0.35	-.02
UCLA Loneliness	257	41.57	10.13	37.51	9.54	50.77***	.20
Child Molester Empathy Measure							
Children in general	139	319.55	59.03	340.86	64.09	19.78***	-.17
Child victim of sexual abuse	139	374.83	71.45	407.35	55.46	42.48***	-.25
Own child victim	139	356.88	89.25	405.91	64.97	71.12***	-.30
Rapist Empathy Measure							
Women in general	69	321.46	72.57	334.54	78.66	2.00	-.09
Woman victim of sexual assault	69	378.88	60.28	409.14	48.84	24.80***	-.27
Own woman victim	69	362.30	83.11	403.99	72.42	32.38***	-.26
Aggression Questionnaire							
Physical aggression	212	17.78	6.72	16.45	6.23	13.81***	.10
Verbal aggression	212	12.66	3.87	12.84	3.83	0.40	-.02
Anger	212	14.27	5.55	13.83	5.23	1.85	.04
Hostility	212	18.67	6.49	17.03	6.39	15.31***	.13
SOARS							
Acceptance of sexual offence	118	25.83	6.00	28.43	4.52	33.04***	-.24
Justifications	118	7.42	4.72	8.31	4.43	3.97*	-.10
Acceptance of offence planning	118	10.64	8.15	17.79	8.96	101.64***	-.39
Acceptance of sexual interests	118	17.80	6.18	19.95	6.98	18.57***	-.16
Acceptance of victim harm	118	26.00	7.00	29.32	4.33	45.46***	-.27
Motivation to change	118	18.84	6.86	20.09	6.67	5.40*	-.09

[†] $p < .10$. * $p < .05$. ** $p < .01$. *** $p < .001$.

DISCUSSION

Almost all of the self-report measures in the NaSOP battery demonstrated good internal consistency in the current sample. Small to medium sized improvement was observed on the dynamic risk assessment instruments and on almost all of the self-report measures over the course of the NaSOP. The improvements on the STABLE-2000 and the VRS:SO suggest that the NaSOP is successfully targeting dynamic risk factors for sexual recidivism. There is evidence that the STABLE-2000 and the VRS:SO are predictive of sexual recidivism (Hanson, 2005; Olver, 2003).

It is less clear, however, how to interpret the improvement on the self-report measures. First, there is the issue of biased responding. Most of the measures were significantly correlated with the BIDR. One interpretation of these findings is that the measures are susceptible to socially desirable responding and, thus, cannot be trusted. An alternative interpretation, however, has recently been presented. Mills and colleagues (Mills & Kroner, 2005, 2006; Mills et al., 2003) have convincingly argued that correlations between a self-report measure and socially desirable responding may occur because both are related to risk and not because the validity of the measure has been compromised by social desirability. Thus, the correlations between social desirability and the measures in the NaSOP battery observed in the current study may not be problematic. A second issue is that most of the self-report measures did not correlate with the risk assessment instruments examined in the current study. The exceptions were the UCLA Loneliness Scale and the Aggression Questionnaire. The UCLA Loneliness Scale was associated with greater risk as measured by the STABLE-2000. The Physical Aggression scale of the Aggression Questionnaire was associated with greater risk as measured by the Static-99 and the SIR-R1.

The finding that self-reported loneliness and physical aggression were correlated with the risk assessment instruments provides support for the concurrent validity of these measures (i.e., that the measures appear to be targeting criminogenic needs). It does not, however, address the predictive validity of these measures (i.e., the extent to which they are related to recidivism). Given that risk assessment instruments do not perfectly predict risk, finding a relationship, or lack thereof, between a self-report measure and a risk measure does not necessarily mean that the measure is similarly related to recidivism. For example, despite the correlation between the UCLA Loneliness scale and the STABLE-2000 found in the current study, researchers have not found the UCLA Loneliness Scale to be predictive of sexual recidivism in prospective studies (Hudson, Wales, Bakker, & Ward, 2002). In contrast, on the Physical Aggression scale of the AG, the findings of the current study and recidivism studies are more consistent. The Aggression Questionnaire's

predecessor, the Buss-Durkee Hostility Inventory (Buss & Durkee, 1957), has been found to predict recidivism in sexual offenders (e.g., Firestone, Nunes, Moulden, & Bradford, 2005). Thus, examining the concurrent validity of the measures in the current study is informative and useful, but does not provide unequivocal evidence concerning the extent to which the measures tap criminogenic needs for sexual offenders. In addition, even for measures of dynamic risk factors with good predictive validity, there is as yet insufficient evidence that *change* on these factors is associated with actual reductions in recidivism (Hanson & Morton-Bourgon, 2005).

Certain limitations of the current research warrant caution in the interpretation of the results. Data on the assessment battery were available for only a fraction of offenders who participated in the NaSOP. It is unknown whether those participants who went unexamined in the current study differ from those who were examined. As a result, the extent to which the current findings are generalizable to all NaSOP participants is unknown. Even for those participants included in the current research, a few of the measures in the NaSOP could not be examined because sample size was far too small. In addition, the smaller sample size did not allow for analyses by offender type for many of the measures. Whenever possible, however, analyses were performed separately for offenders with child victims and those without; in all cases, the results were virtually identical to those reported above.

The reduced sample size also made it impossible to examine the relationship between the pre-post treatment changes and recidivism. The number of offenders in the sample who had been released was too small and, consequently, the recidivists among them were too few for analysis. We have, however, examined recidivism in a partially overlapping sample of NaSOP in a separate research report (Cortoni & Nunes, 2007).

Perhaps a greater threat to validity was the absence of a comparison group. A more rigorous test of the effectiveness of the program would have included a comparison group (e.g., offenders on a waiting list) who completed the same assessment battery at similar time intervals but did not participate in the program. If greater improvement was observed in the NaSOP participants than in the wait-list group, one could be more confident that the improvement was a result of the program. With the design used in the current study, however, the results are open to alternate explanations. For example, it is possible that the improvement observed was simply the result of the passage of time, repeated testing, or some other factor unrelated to the program itself. Future research should include comparison groups and examine the relationship between change on dynamic risk factors and recidivism in sexual offenders (Hudson et al., 2002).

Despite these limitations, the results of the current study are encouraging. As in research on other sex offender programs (e.g., Beech, Fisher, & Beckett, 2004; Eastman, 2004) improvement was observed on a number of areas targeted by the program. Positive change was observed on dynamic risk assessment instruments as well as on self-report measures. In combination with the significant reductions in recidivism associated with participation in the NaSOP (Cortoni & Nunes, 2007), the evidence suggests that the NaSOP is an effective program for sexual offenders.

REFERNCES

Beech, A. R., Fisher, D., & Beckett, R. C. (1999). *STEP 3: An evaluation of The Prison Sex Offender Treatment Programme.* U.K. Home Office Occasional Report retrieved March 20, 2006 from http://www.homeoffice.gov.uk/rds/pdfs/occ-step3.pdf.

Beech, A., Fisher, D., & Bishopp, D. (2004, October). Pre-post change designs. In A. J. R. Harris (Chair), *The future of sex offender treatment outcome research II: The Collaborative Outcome Data committee.* Symposium conducted at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, Albuquerque, NM.

Bumby, K. M. (1996). Assessing the cognitive distortions of child molesters and rapists: Development and validation of the MOLEST and RAPE scales. *Sexual Abuse: A Journal of Research and Treatment, 8,* 37-54.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21*, 343-349.

Buss, A. H., & Perry, M. (1992). The Aggression Questionnaire. *Journal of Personality and Social Psychology, 63,* 452-459.

Cortoni, F., & Nunes, K.L. (2007). Assessing the effectiveness of the National Sexual Offender Program. Research Report R-183. Ottawa, ON: Correctional Service of Canada.

Eastman, B. (2004). Assessing the efficacy of treatment for adolescent sex offenders: A cross-over longitudinal study. *The Prison Journal, 84,* 472-485.

Fernandez, Y. M., Marshall, W. L., Lightbody, S., & O'Sullivan, C. (1999). The Child Molester Empathy Measure: Description and examination of its reliability and validity. *Sexual Abuse: A Journal of Research and Treatment, 11,* 17-31.

Firestone, P., Nunes, K. L., Moulden, H., Broom, I., & Bradford, J. M. (2005). Hostility and recidivism in sexual offenders. *Archives of Sexual Behavior, 34,* 277-283.

Gallagher, C. A., Wilson, D. B., Hirschfield, P., Coggeshall, M. B., & MacKenzie, D. L. (1999). A quantitative review of the effects of sex offender treatment on sexual reoffending. *Correctional Management Quarterly, 3(4),* 19-29.

Hanson, R. K. (2005, June). *The assessment of criminogenic needs of sexual offenders by community supervision officers: Reliability and validity.* Paper presented at the Annual Convention of the Canadian Psychological Association, Montreal, QC.

Hanson, R. K., Gordon, A., Harris, A. J. R., Marques, J. K., Murphy, W., Quinsey, V. L., & Seto, M. D. (2002). First report of the collaborative outcome data project on the effectiveness of

psychological treatment for sex offenders. *Sexual Abuse: A Journal of Research and Treatment, 14,* 169-194.

Hanson, R. K., & Harris, A. J. R. (2000). *STABLE-2000.* Unpublished manuscript. Department of the Solicitor General Canada. Available from the authors: e-mail Andrew Harris at harrisaj@csc-scc.gc.ca.

Hanson, R. K., & Harris, A. J. R. (2001). A structured approach to evaluation change among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 13,* 105-122.

Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis.* Ottawa: Public Safety and Emergency Preparedness Canada.

Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73,* 1154-1163.

Hanson, R.K. & Thornton, D. (1999). *Static 99: Improving actuarial risk assessment for sex offenders.* (User Report No. 99-02). Ottawa: Corrections Research, Department of the Solicitor General of Canada.

Hanson, R. K., & Wallace-Capretta, S. (2004). Predictors of criminal recidivism among male batterers. *Psyhcology, Crime, and Law, 10,* 413-427.

Hudson, S. M., Wales, D. S., Bakker, L., & Ward, T. (2002). Dynamic risk factors: The Kia Marama evaluation. *Sexual Abuse: A Journal of Research and Treatment, 14,* 103-119.

Lösel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1,* 117-146.

Miller, R. S., & Lefcourt, H. M. (1982). The assessment of social intimacy. *Journal of Personality Assessment, 46,* 514-518.

Mills, J. F., & Kroner, D. G. (2005). An investigation into the relationship between socially desirable responding and offender self-report. *Psychological Services, 2,* 70-80.

Mills, J. F., & Kroner, D. G. (2006). Impression management and self-report among violent offenders. *Journal of Interpersonal Violence, 21,* 178-192.

Mills, J. F., Loza, W., & Kroner, D. G. (2003). Predictive validity despite social desirability: Evidence for the robustness of self-report among offenders. *Criminal Behaviour and Mental Health, 13,* 140-150.

Olver, M. E. (2003). *The development and validation of the Violence Risk Scale: Sex Offender version (VRS:SO) and its relationship to psychopathy and treatment attrition.* Unpublished

doctoral dissertation, Department of Psychology, University of Saskatchewan, Saskatoon, SK, Canada.

Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology, 46,* 598-609.

Peacock, E. J. (2000, November). *Measuring readiness for sex offender treatment.* Paper presented at the Annual Research and Treatment Conference of the Association for the Treatment of Sexual Abusers, San Diego, CA.

Rice, M. E., & Harris, G. T. (2003). The size and sign of treatment effects in sex offender therapy. *Annals of the New York Academy of Sciences, 989,* 428-440.

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes: Vol. 1. Measures of social psychological attitudes series (pp. 1-16).* San Diego, CA: Academic Press.

Russell, D., Peplau, L. A., & Cutrona, C. E. (1980). The Revised UCLA Loneliness Scale: Concurrent and discriminant validity evidence. *Journal of Personality and Social Psychology, 39,* 472-480.

Wong, S., Olver, M.E., Nicholaichuk, T.P., & Gordon, A. (2000). The *Violence Risk Scale: Sexual Offender Version* (*VRS:SO*). Regional Psychiatric Centre and University of Saskatchewan: Saskatoon, SK, Canada.