

Changes in J-SOAP-II and SAVRY Scores over the Course of Residential, Cognitive-Behavioral  
Treatment for Adolescent Sexual Offending

Jodi L. Viljoen

Andrew L. Gray

Catherine Shaffer

Simon Fraser University

Natasha E. Latzman

Centers for Disease Control and Prevention

Mario J. Scalora

University of Nebraska-Lincoln

Daniel Ullman

Lincoln Regional Centre

Sexual Abuse: A Journal of Research and Treatment

DOI: <https://doi.org/10.1177/1079063215595404>

## Author Note

Jodi L. Viljoen, Andrew L. Gray, and Catherine Shaffer, Department of Psychology, Simon Fraser University. Natasha E. Latzman, Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention. Mario J. Scalora, Department of Psychology, University of Nebraska-Lincoln. Daniel Ullman, Whitehall Psychiatric Residential Treatment Facility, Lincoln Regional Centre.

This study is funded by the Woods Charitable Fund. The authors have no financial interest in the tools and are not authors of the tools. The authors thank Chris Evans for providing helpful information regarding reliable change indices.

Disclaimer: The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention or the Woods Charitable Fund.

Correspondence regarding this article should be addressed to Jodi Viljoen, Department of Psychology, Simon Fraser University, 8888 University Drive, Burnaby, British Columbia, Canada V5A 1S6. Email: [jviljoen@sfu.ca](mailto:jviljoen@sfu.ca), Telephone: 778-782-8638, Fax: 778-782-3354

## Abstract

Although the Juvenile Sex Offender Assessment Protocol-II (J-SOAP-II) and the Structured Assessment of Violence Risk in Youth (SAVRY) include an emphasis on dynamic, or modifiable factors, there has been little research on dynamic changes on these tools. To help address this gap, we compared admission and discharge scores of 163 adolescents who attended a residential, cognitive-behavioral treatment program for sexual offending. Based on reliable change indices, one-half of youth showed a reliable decrease on the J-SOAP-II Dynamic Risk Total Score and one-third of youth showed a reliable decrease on the SAVRY Dynamic Risk Total Score. Contrary to expectations, decreases in risk factors and increases in protective factors did not predict reduced sexual, violent nonsexual, or any reoffending. In addition, no associations were found between scores on the Psychopathy Checklist: Youth Version and levels of change. Overall, the J-SOAP-II and the SAVRY hold promise in measuring change, but further research is needed.

*Keywords:* dynamic factors, risk assessment, J-SOAP-II, SAVRY, sexual offending

## Changes in J-SOAP-II and SAVRY Scores over the Course of Residential, Cognitive-Behavioral Treatment for Adolescent Sexual Offending

Although adolescents who sexually offend are sometimes assumed to indefinitely pose a high risk to the public, adolescents' risk can change substantially over time. Some youth show reduced risk and desistance from offending as a result of effective interventions (Letourneau et al., 2013; Worling, Litteljohn, & Bookalam, 2010) or developmental maturation (Moffitt, 1993; Monahan, Steinberg, Cauffman, & Mulvey, 2013; Sweeten, Piquero, & Steinberg, 2013). Other youth continue to offend, engaging in increasingly severe forms of offending as they age (Loeber, Farrington, Stouthamer-Loeber, & White, 2008).

Given that risk may fluctuate, many widely used adolescent risk assessment tools, such as the Juvenile Sex Offender Assessment Protocol-II (J-SOAP-II; Prentky & Righthand, 2003) and the Structured Assessment of Violence Risk in Youth (SAVRY; Borum, Bartel, & Forth, 2006), were developed with an emphasis on *dynamic* risk factors (Vincent, Terry, & Maney, 2009). Dynamic risk factors are modifiable factors (e.g., anger management difficulties, limited parental supervision) that may change as a result of intervention, development, or life events. In contrast, historical factors, such as past offending, cannot be undone once they have occurred.

Despite risk assessment tools' stated emphasis on dynamic factors, little research has been conducted on changes in risk. Thus, we examined the ability of the J-SOAP-II and the SAVRY to measure reliable change over the course of treatment (i.e., from admission to discharge) and whether adolescents who improved were less likely to reoffend. We also tested whether adolescents with psychopathic features showed lower levels of improvement during treatment than other adolescents.

### **Use of the J-SOAP-II and the SAVRY to Measure Change**

The J-SOAP-II and the SAVRY are among the most widely used adolescent risk assessment tools (McGrath et al., 2010; Viljoen, McLachlan, & Vincent, 2010). The J-SOAP-II was designed to assess risk for sexual and nonsexual reoffending among adolescents who have sexually offended (Prentky & Righthand, 2003). Although the SAVRY is not designed specifically for adolescents who have sexually offended (Borum et al., 2006), it may be relevant to this population as adolescents who have committed sexual offenses have some similarities to adolescents who have committed nonsexual offenses (e.g., antisocial attitudes and traits; Seto & Lalumière, 2010). Furthermore, adolescents who have committed sexual offenses are more likely to reoffend with nonsexual crimes (e.g., assaults, property crimes) than with sexual crimes (Caldwell, 2010), suggesting that tools like the SAVRY may be useful.

The J-SOAP-II and the SAVRY both include dynamic factors. On the J-SOAP-II, approximately half of the items are purported to be dynamic factors (i.e., 12/28 items = 43%), including all of the items on the Intervention scale (e.g., empathy) and the Community Stability/Adjustment scale (e.g., management of sexual urges). On the SAVRY, approximately two-thirds of the items are dynamic (i.e., 20/30 items = 67%), including items on the Social/Contextual section (e.g., peer delinquency), the Individual/Clinical section (e.g., anger management problems), and the Protective Factors section (e.g., prosocial involvement).

To date, numerous studies have been conducted to examine the predictive validity of the

J-SOAP-II and the SAVRY. Across individual studies, the predictive validity of the J-SOAP-II is mixed (Hempel, Buck, Cima, & van Marle, 2013; Viljoen, Mordell, & Beneteau, 2012b). However, when aggregated across studies, the J-SOAP-II total scores show a moderate ability to predict sexual and nonsexual reoffending (weighted area under the curve [AUC] = .67 and .66 respectively; Viljoen et al., 2012b). Similarly, the SAVRY risk total scores show moderate effect sizes in predictions of violent and general reoffending (weighted  $r = .30$  and  $.32$  respectively; Olver, Stockdale, & Wormith, 2009; see also Singh, Grann, & Fazel, 2011).

The dynamic sections on the J-SOAP-II and the SAVRY have also been found to predict reoffending (Guy, 2008; Viljoen et al., 2012b; Vincent, Chapman, & Cook, 2011). However, we know very little about the ability of the J-SOAP-II and the SAVRY to measure *changes* in reoffense risk. This is because in the vast majority of studies, researchers have administered tools at a single time point. In one of the few studies to examine change, youth showed greater improvements on the J-SOAP-II when the treatment dose was moderate than when it was low or high (Rehfuss et al., 2012). In addition, in a conference presentation, Hilterman (2014) found different trajectories of change on the SAVRY, with some youth increasing and other youth decreasing in risk.

Given that in these two studies researchers focused mainly on the process of change rather than the ability of the J-SOAP-II and the SAVRY to measure change, a couple of key questions remain. First, can the J-SOAP-II and the SAVRY be used to *reliably* measure change? That is, do raters show adequate interrater reliability in assessing change? Second, when can we conclude that a *reliable* change has occurred on these tools? For instance, if an adolescent scores a couple of points lower on the J-SOAP-II or the SAVRY this could simply be due to measurement error, as no tool has perfect reliability. To examine this, we used reliable change indices (RCIs) to estimate reliable or true change after taking into account measurement error (Jacobsen & Traux, 1991). Although many scholars recommend increased use of RCIs (Duff, 2012; Marsden et al., 2011; Stein, Luppá, Brähler, König, & Riedel-Heller, 2010; Wise, 2004), as of yet, few studies have been conducted examining RCIs in the context of violence risk assessment (i.e., Draycott, Kirkpatrick, & Askari, 2012; Kroner & Yessine, 2013; Viljoen et al., 2012a).

### **Changes in Risk Ratings and Reoffending**

If the J-SOAP-II and the SAVRY are able to adequately capture changes in risk, one might expect that decreases in risk scores during treatment are predictive of lower rates of reoffending. Although this research question has not yet been explored in *adolescent* samples, a number of researchers have tested associations between changes in scores on risk assessment tools and reoffending in *adult* offenders. In one of the first studies, Olver and colleagues (2007) found small inverse associations between changes in scores on the Violence Risk Scale–Sexual Offender Version (VRS-SO; Wong, Olver, Nicholaichuk, & Gordon, 2003). This association did not reach significance in the full sample ( $r = -.09$ ) or for the low risk group ( $r = .01$ ), but was significant for high risk offenders ( $r = -.15, p < .05$ ). In other words, high risk offenders who showed greater reductions in risk scores were less likely to reoffend (see also Olver, Nicholaichuk, Kingston, & Wong, 2014). For low risk offenders, the level of improvement may not matter as much as it does for high risk offenders because low risk offenders are already relatively unlikely to reoffend.

In several additional studies with the VRS-SO and other tools (e.g., Violence Risk Scale [Wong & Gordon, 2006]; Level of Service Inventory–Revised [Andrews & Bonta, 1995]), researchers have also reported small inverse correlations between change scores and reoffending (Beggs & Grace, 2011; Labrecque, Smith, Lovins, & Latessa, 2014; Lewis, Olver, & Wong, 2013; Olver et al., 2014; Vose, Lowenkamp, Smith, & Cullen, 2009; Vose, Smith, & Cullen, 2013). In other studies, however, reductions in risk factors have not translated into reduced reoffending (Barnett, Wakeling, Madeville-Norden, & Rakestrow, 2012, 2013; Bowen, Gilchrist, & Beech, 2008; Goodman-Delahunty & O’Brien, in press; Kroner & Yessine, 2013; Serin, Gobeil, & Preston, 2009; Woessner & Schwedler, 2014).

In part, these non-significant findings could be due to methodological issues, such as small sample sizes. Beyond this, the inconsistent results could suggest that some tools do a better job than others at capturing changes related to reoffense risk. To add to this research, the current study is the first to focus on the relationship between changes on *adolescent* risk assessment tools and reoffending. Furthermore, it is one of the few studies in which changes in protective factors were examined.

### **Psychopathic Features and Changes in Risk Ratings**

A final area of focus in the present study is the relationship between change scores and psychopathic features. Psychopathy is a set of traits that is characterized by callousness towards others, limited capacity to experience emotions, and impulsiveness (Skeem, Polaschek, Patrick, & Lilienfeld, 2011). Youth with psychopathic features show higher rates of offending than do other youth (Edens, Campbell, & Weir, 2007). Furthermore, they often show limited treatment compliance (Falkenbach, Poythress, & Heide, 2003; O’Neill, Lidz, & Heilbrun, 2003) and a diminished response to treatment (Manders et al., 2013; O’Neill et al., 2003).

That said, treatment appears to be more effective for these youth than incarceration (Caldwell, Skeem, Salekin, & Van Rybroek, 2006). Also, youth with psychopathic features appear to respond positively to certain forms of treatment (Salekin, Worley, & Grimes 2010). Multisystemic Therapy, for instance, is associated with significant decreases in parent-reported psychopathic features (Butler, Baruch, Hickey, & Fonagy, 2011). Functional Family Therapy has been found to result in improved behavioral, emotional, and social adjustment in youth with callous-unemotional features (White, Frick, Lawing, & Bauer, 2012). In another study, adolescents with psychopathic features responded to a brief 12 session intervention that focused on motivational and cognitive-behavioral elements (Salekin, Tippey, & Allen, 2012).

Typically, researchers have measured the success of treatment via an examination of reoffense rates or changes in symptoms of psychopathy or conduct disorder (e.g., Butler et al., 2011; White et al., 2012). However, risk assessment tools may also provide a useful indicator of treatment-related improvement (see Olver, Lewis, & Wong, 2013). Thus, in the present study, we examined whether *adolescents* with psychopathic features demonstrate fewer reductions in risk factors and gains in protective factors during treatment than other adolescents.

### **Present Study**

Although the J-SOAP-II and the SAVRY were designed to measure change, there is, as of yet, little research on their ability to do so. Thus, we examined the interrater reliability of

ratings of change on the J-SOAP-II and the SAVRY; the proportion of youth who showed reliable change in J-SOAP-II and the SAVRY risk scores during residential cognitive-behavioral treatment (CBT); and whether improvements (i.e., reduced risk scores and increased protective scores) were associated with lower reoffense rates. In addition, we examined whether youth with psychopathic features were less likely to show treatment-related improvements.

It was predicted that the J-SOAP-II and the SAVRY would show adequate reliability for measuring change. Given that CBT is associated with significant reductions in sexual reoffending when compared to treatment as usual (odds ratio = 0.59; Reintzel & Carbonell, 2006), it was hypothesized that adolescents attending the program would show reductions in risk scores and increases in protective factors from admission to discharge. Consistent with adult studies, it was expected that, after controlling for risk level, adolescents who improved would be less likely to commit sexual and nonsexual reoffenses. Finally, it was hypothesized that adolescents low in psychopathic features would show more improvement than those high in these features.

## Method

### Participants

Potential participants included all of the 169 male adolescents who were discharged between January 1993 and December 2004 from a nonsecure residential sex offender treatment program in a medium-sized, mid-Western American city, namely the Whitehall Psychiatric Residential Program in Lincoln, Nebraska. To be included in the present study, youth had to have remained in the program for a sufficiently long enough period of time (i.e., 30 days or more) that they had an opportunity to show change on the risk assessment tools. Six youth were omitted because they were discharged in less than 30 days after admission, resulting in a sample size of 163.

The mean age of the youth at the time of admission was 15.39 years ( $SD = 1.50$ ). Although a large majority of the youth were non-Hispanic Caucasian (82.8%,  $n = 135$ ), a small proportion were African American (8.6%,  $n = 14$ ), Hispanic (4.9%,  $n = 8$ ), American Indian or Alaskan Native (1.2%,  $n = 2$ ), or biracial (2.5%,  $n = 4$ ). The length of time youth spent in the program ranged from 31 days to 4.07 years; the mean number of years in the program was 1.13 ( $SD = 0.67$ ). Youth had committed a variety of sexually abusive behaviors (i.e., index offenses) that led to treatment, including genital penetration (36.8%,  $n = 60$ ), anal penetration (35.0%,  $n = 57$ ), oral-genital contact (48.5%,  $n = 79$ ), fondling (62.0%,  $n = 101$ ), and exhibitionism (14.1 %,  $n = 23$ ).

In most cases, youth had at least one index offense victim who was three or more years younger than the youth (86.7 %,  $n = 137$ ). Approximately half of the youth had index offenses against female-only victims (46.3%,  $n = 74$ ), 24.4 % ( $n = 39$ ) had male-only victims, and 29.4% ( $n = 47$ ) had both female and male victims. Many of the victims were related to the offender (71.9%,  $n = 115$ ). Approximately half of the youth had committed prior sexual offenses (50.9%,  $n = 83$ ) or were charged or convicted for nonsexual offenses (51.0%,  $n = 80$ ).

This sample of youth has been included in previous research on risk and protective factors (Elkovitch, Viljoen, Scalora, & Ullman, 2008; Latzman, Viljoen, Scalora, & Ullman,

2011; Spice et al., 2013; Viljoen, Elkovitch, Scalora, & Ullman, 2009; Viljoen et al., 2008). However, the current study has a different focus (i.e., dynamic change) and does not include any analyses that are redundant with prior work.

### Procedure

Ethics approval was obtained from the University of Nebraska-Lincoln, Simon Fraser University, and the research site. This study had a quasi-prospective design with the risk assessments being made in the context of research rather than clinical practice. Three trained research assistants rated the J-SOAP-II and the SAVRY for each youth based on archival file information. Youth's admission ratings on these tools were made using the file information available at admission, and youth's discharge ratings were made using the file information available at discharge. After all J-SOAP-II and SAVRY ratings were completed, two different research assistants separately completed ratings on the Hare Psychopathy Checklist: Youth Version (PCL:YV; Forth, Kosson, & Hare, 2003). By rating the PCL:YV separately, it minimized the possibility that PCL:YV ratings might influence assessments of how much a youths' risk and protective factors changed.

Research assistants followed the rating guidelines in the manuals for the J-SOAP-II, the SAVRY and the PCL:YV; no adaptations or changes were made to any rating criteria. In coding the measures, research assistants were blind to youths' subsequent charges and convictions. All research assistants were Ph.D. students in clinical forensic psychology, had completed graduate coursework on risk assessment, and had been employed in clinical practicum positions with offenders. Prior to commencing coding, raters underwent didactic training, received readings, and completed five practice cases with the study measures using case files.

Given that the youth in this sample had, on average, spent approximately one year in the residential treatment program, the file information available to code the study measures was comprehensive. On average, files were over 600 pages in length and included psychiatric assessments, psychological assessments, nursing records, medical examination information, social work reports, teacher assessments, school records, treatment plans and records, progress notes, physician orders, arrest records, and other materials. Raters coded the quality of each file on a scale of 1 to 10, with 1 being *extremely poor in quality* and 10 being *extremely good in quality*. The modal quality rating was generally good ( $Mode = 8.00$ ,  $Mdn = 7.00$ ,  $M = 7.29$ ,  $SD = 1.31$ ), with only five files receiving scores of five or less. As the files generally contained the necessary information for coding tools, missing data were scarce; no youth were missing data on the J-SOAP-II and only one youth was missing information for items on the SAVRY (this case was prorated for the one missing item at admission and four missing items at discharge). To examine interrater reliability of the risk assessment tools, a random sample of files (22.7%,  $n = 37$ ) were selected and separately coded by a second rater. As described in the Results section, interrater reliability was generally good to excellent.

Approximately 1.37 years after completing the coding of the study measures, youths' post-discharge juvenile justice and adult criminal records were obtained through statewide law enforcement and probation. Records were available for all participants in our sample. The average length of the post-discharge follow-up period was 8.07 years ( $SD = 3.50$ ), but ranged from 2.18 years to 13.56 years as youth were discharged at different dates. During the follow-up period, 7.4% of youth were arrested for sexual reoffenses ( $n = 12$ ), 12.9% for violent nonsexual

reoffenses such as assault ( $n = 21$ ), and 46.0% for any reoffense ( $n = 75$ ). This latter category included property offenses, violent nonsexual offenses, sexual offenses, and miscellaneous offenses (e.g., mischief) but did not include traffic offenses (e.g., speeding tickets). A reoffense was defined as an arrest rather than conviction, as sexual offenses are sometimes reduced to nonsexual offenses through plea bargains (Letourneau, Armstrong, Bandyopadhyay, & Sinha, 2013). To ensure a consistent and transparent reporting of methodology and results, this manuscript adheres to the Risk Assessment Guidelines for the Evaluation of Efficacy (RAGEE) Statement (Singh, Yang, Mulvey, & the RAGEE Group, 2014), a 50-item reporting checklist.

### **Description of Treatment Program**

The Whitehall Psychiatric Residential Treatment Program is a specialized, community-based residential program that provides treatment to youth adjudicated for a sexual offense. To be admitted into the program, youth had to meet the following admission criteria: between 13 and 17 years of age, intellectual and adaptive functioning at least at the borderline level, adjudicated of a sexual offense and mandated to receive treatment, and demonstrated self-control that would allow functioning in an open, unlocked treatment program. The program is staffed by a multidisciplinary team including a licensed clinical psychologist, a psychiatrist, Masters-level mental health clinicians, nurses, occupational therapists, recreational therapists, and Bachelors-level direct care staff.

At intake, all youth undergo a comprehensive psychological evaluation and throughout treatment both youth and treatment staff complete on-going assessments that assess behavioral and emotional symptomatology. Treatment plans are individualized to meet each youth's strengths and treatment needs, although they tend to focus on similar themes: insight and accountability for past offenses, problem-solving skills, skill building and promotion of positive relationships, development of relapse-prevention plans, enhancing awareness of victim impact, reduction of psychopathology, and educational success. These areas are addressed via several modalities including individual, group, and family therapy, as well as school-based interventions and recreational and occupational therapy. Youth attended individual therapy two to five times a week, as well as a relapse prevention group (three times a week), occupational therapy, and recreational therapy. Depending on the youth's needs, they also attended trauma-focused, coping skills, and relationship skills groups all with a CBT, skill-building orientation.

The Whitehall Program is an unlocked program. The daytime level of supervision is comparable to a day treatment program and youth have more community contact than a traditional secure or correctional facility. For instance, depending on their progress in the program, youth can go home with their family for a weekend or go on a community outing (e.g., out for lunch with their family). Also, at the time period captured by the current study, some youth in the program were attending public schools and/or church in the community.

### **Measures**

**Juvenile Sex Offender Assessment Protocol-II (J-SOAP-II).** The J-SOAP-II (Prentky & Righthand, 2003) is a 28-item checklist designed to aid in assessing risk for sexual violence and general delinquency. It is intended for use with adolescents, aged 12 to 18, who have a history of sexually coercive behavior. In the present study, we focused on the Intervention and Community Stability/Adjustment scales, as the J-SOAP-II authors conceptualize these scales as

dynamic. These scales contain seven and five items, respectively, which are rated on a three-point scale (*absent, possibly present, clearly present*) and are summed to create a Dynamic Risk Total Score. The J-SOAP-II does not have cut-off scores or yield probability estimates.

In the J-SOAP-II manual, the authors state to omit the Community Stability/Adjustment scale if a youth is “incarcerated in a correctional facility or a secure residential treatment program” (p. 25). However, this scale can be rated for youth in nonsecure residential settings (Prentky et al., 2010). The residential treatment program in this study was nonsecure and unlocked. For instance, youth in the program had numerous outings in the general community, such as home visits and attendance at school and church. Thus, similar to Prentky et al. (2010), we rated this scale for the youth in our study.

A meta-analysis indicated that the J-SOAP-II’s Intervention and Community Stability/Adjustment scales significantly predicted sexual and nonsexual reoffending with small to moderate effect sizes (Viljoen et al., 2012b). In prior studies, researchers have found the Intervention scale to have good to excellent interrater reliability, and the Community Stability/Adjustment scale to have fair to excellent interrater reliability (e.g., Aebi, Plattner, Steinhausen, & Bessler, 2011; Caldwell, Ziemke, & Vitacco, 2008; Martinez, Florez, & Rosenfeld, 2007; Rajlic & Gretton, 2010). In the present study, internal consistency was adequate (alphas > .77; see Table 1) except for the Community Stability/Adjustment scale at admission (alpha = .60).

**Structured Assessment of Violence Risk in Youth (SAVRY).** The SAVRY (Borum et al., 2006) is a 30-item checklist that was designed to assess violence risk in male and female adolescents. The SAVRY is based on a structured professional judgment (SPJ) model and does not have cut-off scores. In the present study, we focused on the Social/Contextual, Individual/Clinical, and Protective Factors sections, as the SAVRY authors conceptualize these sections as dynamic. The Social/Contextual and Individual/Clinical sections contain six and eight items, respectively, which are rated on a three-point scale (with ratings of *low, moderate, or high risk*). The Protective Factors section contains six items, which are rated dichotomously (*present or absent*). Consistent with other research on the SAVRY (e.g., Lodewijks, de Ruiter, & Doreleijers, 2010), we summed items to form scores for each section and created a Dynamic Risk Total Score by summing scores on the Social/Contextual and Individual/Clinical sections.

In a meta-analysis examining the predictive validity of SPJ based risk assessment tools, Guy (2008) reported that the SAVRY Social/Contextual, Individual/Clinical, and Protective Factors sections significantly predicted physical and sexual violence and nonviolent reoffending, with weighted AUC scores ranging from .64 to .75 (see also Lodewijks, et al., 2010; Vincent, Guy, Gershenson, & McCabe, 2012). Researchers have found these sections to have good to excellent interrater reliability (e.g., intraclass correlation coefficients > .80; Lodewijks, Doreleijers, de Ruiter, & Borum, 2008). Internal consistency was generally acceptable in the present study but was low for the Social/Contextual and Protective Factors sections (alphas < .70; see Table 1).

**Hare Psychopathy Checklist: Youth Version (PCL:YV).** The PCL:YV (Forth et al., 2003) is a 20-item rating scale designed to measure psychopathic traits. This measure was adapted for adolescents from the Hare Psychopathy Checklist – Revised (PCL-R; Hare, 1991,

2003). Each PCL:YV item is rated on a 3-point scale (i.e., 0, 1, 2), with higher scores indicating a larger number of psychopathy-related traits. Consistent with the PCL:YV manual, items were summed to form a PCL:YV Total Score and scores on four facets. The Interpersonal facet includes four items (e.g., grandiose sense of self-worth), the Affective facet includes five items (e.g., callous/lacking empathy), the Behavioral facet includes five items (e.g., irresponsibility), and the Antisocial facet includes five items (e.g., early behavior problems).

In prior studies, researchers have found that the PCL:YV is a valid and reliable measure available for assessing psychopathic features (Edens et al., 2007; Salekin et al., 2004). Furthermore, the PCL:YV was found to be a significant predictor of some forms of reoffending in a sample of adolescents who sexually offended (Gretton et al., 2001). Similar to other studies (Forth et al., 2003; O'Neill et al., 2003), internal consistency in the present study was acceptable for the total score ( $\alpha = .80$ ). However, it was modest for the facet scores (alphas = .59, .64, .51, and .68 for Interpersonal, Affective, Behavioral, and Antisocial facets, respectively) possibly due to the small number of items in each facet (i.e., 4 or 5 items each; see Cortina, 1993). Consistent with previous research (e.g., Spain, Douglas, Poythress & Epstein, 2004), the interrater reliability of the PCL:YV total and facet scores generally fell in the excellent range (intraclass correlation coefficients for single raters, absolute agreement, two-way random effects model = .89 for total score, and .83, .89, .68, and .85 for Interpersonal, Affective, Behavioral, and Antisocial facets, respectively, based on a random sample of 25 cases from the present study). The mean PCL:YV score was 17.25 ( $SD = 6.00$ ). This is consistent with prior research with samples of adolescents in residential treatment programs (e.g., Marshall, Egan, English, & Jones, 2006), but slightly lower than other samples of youth who have sexually offended (e.g., Gretton et al., 2001).

### Data Analysis

To examine interrater reliability of ratings of change on the J-SOAP-II and the SAVRY, intraclass correlation coefficients (ICCs) were calculated on a random sample of 37 cases (22.7%). We used a random effects model for single raters and examined absolute agreement rather than general consistency (McGraw & Wong, 1996). ICCs are commonly classified in the following manner (Cicchetti & Sparrow, 1981; Shrout & Fleiss, 1979): poor ( $\leq .40$ ), fair (.40 to .59), good (.60 to .74), and excellent ( $\geq .75$ ).

To examine level of change in J-SOAP-II and the SAVRY scores from admission to discharge, repeated-measures multivariate analyses of variance (MANOVA) were conducted using the Dynamic Risk Total Scores and scale scores. Magnitude of change was interpreted based on Cohen's  $d$  for repeated measures, where .20 corresponds to a small effect size, .50 to a medium effect size, and .80 to a large effect size (Cohen, 1988). To determine the proportion of adolescents who showed reliable increases or decreases in scores, we calculated reliable change indices (RCIs, 95% confidence intervals) with the Jacobson and Truax (1991) formula.<sup>1</sup> The RCI takes into account measurement error by calculating whether an individual showed more change than would be expected based on chance or error alone. Whereas group analyses such as  $t$ -tests can sometimes mask individual changes (e.g., if an equal proportion of youth increase and decrease, these effects could cancel each other out, resulting in a non-significant  $t$ -value), RCIs provide *individual-oriented* analyses by examining the proportion of individuals who show reliable increases, reliable decreases, and no reliable change in scores. Although reliable change

can be calculated with various forms of reliability (Evans, Margison, & Barkham, 1998; C. Evans, personal communication, September 18, 2014), in this study, RCIs were calculated based on interrater reliability ratings (i.e., ICCs at admission). This is because interrater reliability is a critical form of reliability for risk assessment tools, as risk assessment tools require rater judgment. For comparison, we also calculated RCIs based on internal consistency.

To analyze the association between change and reoffending, change scores were calculated for each scale as follows: [Change Scores for Risk Scales = Score at Admission minus Score at Discharge] and [Change Score for Protective Scale = Score at Discharge minus Score at Admission]. Thus, higher change scores indicated greater improvements. Consistent with research on adult tools (e.g., Olver et al., 2014), we examined zero-order correlations between change scores and reoffending and then conducted three sets of partial correlations controlling for 1) static risk level (i.e., J-SOAP-II Static Scale for J-SOAP-II analyses and SAVRY Historical Factors Section for SAVRY analyses), 2) admission score on the respective scale, and 3) treatment length. To determine whether the presence of reliable change added incrementally to the prediction of reoffending relative to static risk, a series of logistic regression analyses were conducted. Given the modest base rates for sexual and violent nonsexual reoffending, penalized likelihood regression was conducted to reduce the risk of bias in the estimation of the odds ratio (Heinze, 2006). Although penalized likelihood methods may be applied to Cox regression, logistic regression remained the preferred method of analysis as the exact dates of reoffense could not be ascertained for all youth ( $n = 6$ ).<sup>2</sup> Finally, receiver operating characteristic (ROC) curve analyses were conducted to generate the area under the curve (AUC) values for J-SOAP-II and SAVRY scores at admission and discharge (Hanley & McNiel, 1982). Comparative analyses between the admission and discharge AUC values were conducted using the method developed by DeLong, DeLong, and Clarke-Pearson (1988).

To analyze the association between psychopathy scores and reoffending, zero-order correlations were calculated between PCL:YV total and facet scores and reoffending outcomes. Furthermore, given that associations between psychopathic features and change scores may not be notable unless youth reach a certain threshold of these features, we compared youth scoring high, moderate, and low on the PCL:YV. Cut-offs were selected based on quartiles; youth who scored at 25<sup>th</sup> percentile or lower (i.e.,  $\leq 13$ ) were classified as low, those who scored between the 25<sup>th</sup> and 75<sup>th</sup> percentiles were classified as moderate (i.e., 14 to 21), and those who scored at the 75<sup>th</sup> percentile or higher (i.e.,  $\geq 22$ ) were classified as high.

Analyses were generally conducted in IBM SPSS, Version 19. However, AUCs and penalized likelihood regression were performed in R (Heinze & Ploner, 2004; Robin et al., 2011), which has increased capacities for these analyses (e.g., R provides the DeLong et al.[1988] test; see R Core Team, 2014). All  $p$ -values for analyses were set at  $p < .05$ , and family-wise corrections were made where applicable.

## Results

### Reliability of Ratings of Change

For the J-SOAP-II, ICCs for change scores were good to excellent for each scale (.64 to .82), indicating that change on the J-SOAP-II can be measured with adequate interrater reliability (see Table 1). On the SAVRY, ICCs for change scores were good for the Individual/Clinical

section and Dynamic Risk Total Score (.71 and .66, respectively), but fair for the Social/Contextual section (.46) and poor for the Protective Factors section (.24). As shown in Table 1, interrater reliability at discharge was generally higher than at admission.

### **Level of Change from Pre- to Post-Treatment**

Based on a repeated-measures multivariate analysis of variance (MANOVA), significant multivariate effects were found across the within-subjects time points (i.e., admission and discharge) for both the J-SOAP-II dynamic risk scales (Phillai's Trace  $[V] = 0.65$ ,  $F(2, 161) = 147.60$ ,  $p < .001$ ) and SAVRY dynamic risk scales (Phillai's Trace  $[V] = 0.63$ ,  $F(3, 160) = 88.87$ ,  $p < .001$ ).<sup>3</sup> Univariate analyses (Table 2) revealed significant decreases from admission to discharge among each of the risk scales with large repeated-measures Cohen's  $d$  effect sizes ( $> .80$ ) for five of the six risk scales (the exception being the Social/Contextual section of the SAVRY which produced a moderate effect size). Furthermore, scores on the SAVRY Protective Factors section significantly increased from admission to discharge; however, the magnitude of the difference was small. Stability coefficients ranged from .62 to .75 (Table 2).

In general, RCI values classified a sizable number of youth as having exhibited a reliable change between admission and discharge. When RCIs were calculated based on interrater reliability, youth needed to show a change of at least 8 points on the J-SOAP-II and the SAVRY Dynamic Risk Total Scores in order for change to be classified as reliable (see Table 1). When RCIs were calculated based on internal consistency, a narrower scope of change was needed to classify it as reliable change, as the J-SOAP-II's and the SAVRY's alphas were typically higher than their ICCs.

Approximately one-half of youth showed reliable decreases on the J-SOAP-II Dynamic Risk Total Scores (see Table 3). Somewhat fewer youth (approximately one-third) showed reliable change on the SAVRY Dynamic Risk Total Score. On the section and scale scores, a relatively high proportion of youth showed reliable decreases on the J-SOAP-II Intervention scale and the SAVRY Individual/Clinical section (38.7 – 50.3%), whereas rates were more modest for the J-SOAP-II Community Stability/Adjustment scale and the SAVRY Social/Contextual sections (6.7 – 19.0%). Although there were no reliable *increases* in risk factors, a sizable proportion of youth did not meet the threshold for reliable change regardless of direction ( $\geq 42.3\%$  per scale). On the SAVRY Protective Factors section, only 8.0% of youth displayed reliable change.

### **Changes in Risk Ratings and Reoffending**

Prior to the main analyses, point-biserial correlations and AUC values were calculated for the admission and discharge scores for the J-SOAP-II and the SAVRY with the reoffense outcomes to determine whether discharge scores were more predictive of reoffending than admission scores (Tables 4 and 5, respectively). Associations between the dynamic risk and protective scores and reoffending were modest, with only a single AUC value being considered moderate in size (i.e.,  $AUC \geq .64$ ; Rice & Harris, 2005). Contrary to expectations, several of the admission scores were stronger predictors of reoffending when compared to their respective discharge scores. However, none of these differences achieved statistical significance using the comparative methods developed by DeLong et al. (1988).

Next, correlational analyses were conducted to examine whether decreased risk factors and increased protective factors predicted lower rates of reoffending (Table 6). The correlations were modest and none reached statistical significance even after controlling for static risk level, admission scores on the respective scale, and treatment length. To examine whether change scores predict reoffending over shorter periods of time (as compared to our average follow-up of 8.07 years), post-hoc correlational analyses were conducted using fixed follow-up periods of 1 and 2 years.<sup>4</sup> For this analysis we controlled for risk level and scores at admission. Given that base rates of reoffending were low for the 1 and 2 year follow-ups (i.e., 2.5% to 3.1%), these analyses focused on the any reoffending outcome which had base rates of 9.9% and 14.9% at 1 and 2 years, respectively. Again, none of the partial correlation coefficients between change scores and reoffending reached significance.

Although associations with change scores were non-significant, we next tested whether reoffending might be inversely associated with *reliable* change (i.e., change that met the threshold to conclude it was reliable rather than measurement error). These results are presented in Tables 7 and 8. Maximum likelihood logistic regression analyses were conducted for the outcome of any reoffending. Penalized logistic regression, using R, was employed for sexual and nonsexual violent reoffending as base rates were modest for these outcomes (i.e., 7.4% and 12.9% respectively; see King & Zeng, 2001). These analyses controlled for static risk level in Step 1. Overall, model fit was poor, and reliable change failed to significantly predict reoffending with several exceptions. First, youth who showed reliable decreases in the Intervention subscale were less likely to sexually reoffend (odds ratio = 0.14,  $p = .013$ ; see Table 7). Second, and in contrast, youth who showed reliable decreases on the J-SOAP-II Community Stability/Adjustment subscale were at *increased* likelihood for sexual reoffending (odds ratio = 6.58,  $p = .022$ ; see Table 7) and any reoffending (odds ratio = 3.06,  $p = .021$ ; see Table 8). Overall, the presence of reliable change failed to add significant incremental validity relative to static risk level for the majority of the analyses; the only exception being two analyses with the J-SOAP-II ( $\Delta\chi^2[2] = 6.88$ ,  $p = .032$ , for any reoffending;  $\Delta\chi^2[2] = 7.64$ ,  $p = .022$  for sexual reoffending).

### Psychopathic Features and Changes in Risk Ratings

None of the correlations between PC:YV total and facet scores were significantly correlated with change, although the correlations were in the anticipated direction (i.e., inverse correlations; Table 9). Similarly, when a MANOVA was conducted (see Table 10), the multivariate effect of PCL:YV groups on change failed to reach significance, and none of the univariate effects were significant either (family-wise error rates were controlled using the Bonferroni correction, i.e.,  $p \leq .05 = .010$ ). We reran analyses using different cut-off scores for psychopathy (i.e., low = scores of  $\leq 15$ , moderate = scores of 16 to 24, large = scores of  $\geq 25$ ), again finding no significant differences. Finally, as evidenced by chi-square analyses (see last column in Table 10), there were no significant associations found between PCL:YV group and rates of *reliable* change on the dynamic scales.

### Discussion

Adolescent risk assessment tools, such as the J-SOAP-II and the SAVRY, include an emphasis on dynamic factors. However, as of yet, little research has been conducted on dynamic changes in these factors. To help address this gap, we compared admission and discharge scores

on the J-SOAP-II and the SAVRY in a sample of 163 adolescents who had participated in a residential CBT treatment program for adolescents who had sexually offended.

### Primary Findings

Adolescents showed substantial changes in their risk ratings from admission to discharge. On the J-SOAP-II, effect sizes for change at an overall group level were large. In addition, one-half of youth showed a reliable decrease on the J-SOAP-II Dynamic Risk Total Score. Although the treatment program was a specialized program targeted at sex offending, youth in the program also showed moderate reductions in general risk factors for violence on the SAVRY. Specifically, one-third of youth showed a reliable decrease on the SAVRY Dynamic Risk Total Score.

Changes in SAVRY Protective Factors were modest in comparison, with only 8% of youth showing a *reliable* increase in protective factors. This could be because treatment programs for sexual offenders generally focus on risk reduction rather than strengths promotion (Ward 2002; Ward & Brown, 2004). Alternatively, the protective factors section of the SAVRY may be less dynamic in nature. For instance, the SAVRY protective factor, *resilient personality traits*, is defined to include “above-average intellectual ability” (Borum et al., 2006, p. 54), which is difficult to modify. Another possibility is that the Protective Factors section is less sensitive to detecting change than the risk scales because it rates items dichotomously (*present* or *absent*) rather than on a three-point scale. Finally, because the Protective Factors section had modest reliability (alpha = .58, ICC = .68), a higher change score was required to conclude that a change was reliable.

Despite the significant changes in youth’s risk ratings from admission to discharge, risk ratings at discharge were no more accurate in predicting reoffending than risk ratings at admission. Although many youth showed improvement over the course of treatment, this generally did not directly translate into reductions in reoffending. One exception to this was that reliable decreases in risk factors on the J-SOAP-II Intervention scale significantly predicted lower rates of sexual reoffending.

The general failure to find associations between change scores and reoffending could indicate that the J-SOAP-II and the SAVRY dynamic scales may not be tapping into all of the relevant dynamic factors. Not only were change scores non-predictive, even the admission and discharge scores on the dynamic scales did not significantly predict reoffending in this sample, although they have shown adequate predictive validity in other studies (Guy, 2008; Viljoen et al., 2012b). Beyond the possibility that these findings may reflect on the tools themselves, these null results may be due to a number of equally plausible or more plausible explanations, such as methodological limitations (e.g., the reliance on official records to measure change) or challenges in sustaining treatment effects.

In particular, if adolescents’ risk is changeable it may not make sense to presume that decreases in risk would predict reduced reoffending eight years later, as youth may have experienced many changes in risk and protective factors during this time (e.g., gains in impulse control with maturation or increased antisocial attitudes with cumulative exposure to antisocial lifestyles). However, in the current study, change scores did not predict reoffending at one-year and two-year fixed follow-ups either. In future research, researchers should test shorter time

intervals (e.g., six-month follow-ups) to determine whether the relevance of change may expire at an even earlier date. The period of transition from residential programs to home environments may be a period of particular fluidity in risk; youth may not necessarily maintain treatment gains as they transition from residential treatment to the community (Nickerson et al., 2007). In particular, given that some adolescents' home environments may be characterized by high levels of conflict and limited supervision (Burns, Hoagwood, & Mrazek, 1999), risk scores may increase after youth return home.

Surprisingly, in the present study, reliable decreases in risk factors on the J-SOAP-II Community Stability/Adjustment subscale were associated with *higher* rates of sexual and any reoffending. This finding is difficult to explain, especially as this scale includes well-established risk factors such as management of anger, management of sexual urges, stability in school, and evidence of positive support systems. However, there are three potential explanations for this: 1) youth showing decreased risk in this domain may have been subject to increased monitoring leading to a higher likelihood of detection; 2), youth who were perceived to have high levels of stability and community adjustment were provided with less supervision upon discharge, increasing their opportunities to reoffend; or 3) youth showed decreased risk in this domain because the treatment program provided a high degree of structure and supervision (e.g., an onsite school). Youth who responded well to this structure may have been vulnerable to relapse (e.g., reoffending) when discharged back into unstructured home environments. Finally, although the manual states it is acceptable to use this scale with youth in a nonsecure residential setting (Prentky & Righthand, p. 25 and 26), and this has been done in prior research (Prentky et al., 2010), it is possible that youths' discharge ratings may provide an unrealistically high estimate of a youth's capacities in these areas. Instead, it is important to not only assess youth during the treatment program, but also to reassess them after they return to their home environment.

Whereas in previous studies researchers have found that youth with psychopathic features are less responsive to treatment than other youth (Manders et al., 2013; O'Neill et al., 2003), no significant differences emerged in the present study. Youth high in psychopathic features appeared to show similar decreases in risk factors and increases in protective factors as other youth. This may be because the residential CBT treatment program that the youth received was appropriate for youth with psychopathic features; there is some evidence that youth with psychopathic features respond quite favorably to some intensive, residential interventions (Caldwell et al., 2006) and cognitive behavioral approaches (Salekin et al., 2012). Another possibility is that reduced response to treatment is only seen in youth with very high levels of psychopathic features, whereas most youth in our sample had mid-range scores on the PCL:YV ( $M = 17.25$ ). Finally, most studies have examined changes in features of psychopathy and conduct disorder as treatment outcomes, whereas the current study focused on risk and protective factors. Thus, it may be that risk and protective factors are more dynamic than psychopathic features. If this is the case, it may be useful to target risk and protective factors in treatment for youth with psychopathic features rather than solely focusing on the reduction of psychopathic features themselves (see Wong & Hare, 2005).

### **Study Limitations**

In interpreting these study results, several caveats are important. First, similar to other

studies on dynamic change (e.g., Olver et al., 2007), the J-SOAP-II and the SAVRY were rated based on file information. Although file- and interview-based ratings are strongly correlated (Gretton et al., 2001), it is possible that demand characteristics impacted ratings (e.g., raters may have rated discharge risk scores lower than warranted). Similar to other risk assessment studies (e.g., Douglas, Ogloff, Nicholls, & Grant, 1993), we coded some files that predated the development of the tools to ensure an adequate sample size and a sufficiently long follow-up period. This means that the files did not necessarily contain specific information that mapped exactly onto the J-SOAP-II and SAVRY factors. Nonetheless, the files were comprehensive in nature, raters judged most files to be high quality, and there was very little missing data (i.e., only one youth had any missing items on the J-SOAP-II and the SAVRY).

Second, similar to other studies (see Viljoen et al., 2012 for a summary), official records were used to measure reoffending. This approach may fail to detect some sexual offenses (Fleming, Jory, & Burton, 2002). Thus, future research should assess reoffending through multiple methods (e.g., youth and parent self-report, treatment records).

Third, the treatment program examined in this study has not previously been researched. Thus, if the results had indicated that youth did not change, this would have been difficult to interpret; such a finding could have meant that the treatment program was ineffective and/or that the tools were not adequately sensitive to change. As it turned out, youth showed significant improvements over the course of treatment. However, without a control group it is not possible to determine if changes in risk scores occurred as a result of treatment and/or other mechanisms (e.g., maturation, regression to the mean).

Fourth, this study focused on the J-SOAP-II scales and the SAVRY sections that the authors conceptualize as dynamic, historical factors should be examined in future work, given the possibility that some of these factors may change over time (e.g., a youth can engage in additional acts of violence or experience maltreatment).

Fifth, although the overall sample size was 163, interrater reliability data was collected for a relatively small subset of these youth (22.7%,  $n = 37$ ); this limits our ability to make firm conclusions about the interrater reliability of change scores. Finally, we did not record information on where youth were residing prior to admission. It is possible that a small number of youth were residing in locked settings prior to admission; staff at the treatment program indicated that such cases would be rare. Also, raters were instructed to follow the J-SOAP-II manual, which states that if a youth was recently in a correctional facility or a secure residential treatment program for longer than 6 months, he must have been in the community for “at least 3 months” in order to rate the Community Stability/Adjustment scale (Prentky & Righthand, 2005, p .25).

## **Implications**

Results of this study have several implications for research and practice. In particular, the finding that adolescents’ risk showed substantial change over the course of treatment reinforces that clinicians should reassess risk regularly. Further research should clarify the optimal interval for reassessment. At the present time, experts recommend reassessing risk at least every six months and at periods of significant change, such as if a youth acquires a new charge or is released from a custodial facility (Vincent, Grisso, & Guy, 2012).

In addition, given that the J-SOAP-II and the SAVRY detected relatively high rates of change and generally showed adequate interrater reliability in measuring change, these tools hold promise as measures of changes in risk and protective factors. However, to determine if certain approaches are more sensitive to change than others, researchers should compare these and other approaches for measuring change (e.g., the Violence Risk Scale: Youth Version [Wong, Lewis, Stockdale, & Gordon, 2011], the Estimate of Risk of Adolescent Sex Offense Recidivism [ERASOR; Worling & Curwen, 2001], the Short-Term Assessment of Risk and Treatability: Adolescent Version [Viljoen, Nicholls, Cruise, Desmarais, & Webster, 2014]). Rather than focusing on the predictive validity of change scores (and conceptualizing absence of change as another risk factor), researchers should also examine the extent to which measuring change can guide refinements to treatment plans.

Finally, the results of this study indicate that clinicians and researchers should use caution in interpreting change. If a youth's score changes by a couple of points on a tool, it does not mean that he or she showed meaningful change, as all tools have a certain degree of imprecision. Indeed, we found that a youth's score on the J-SOAP-II and the SAVRY Dynamic Risk Total scores had to have increased or decreased 8 points in order to conclude that a youth had shown reliable change (after taking into account imperfect interrater reliability). To guide the interpretation of changes in risk, test developers and researchers could provide RCIs or other empirically-derived guidelines. Other types of clinical measures, such as treatment outcome measures (Lambert et al., 1996) and neuropsychological tests (Strauss, Sherman, & Spreen, 2006), provide this type of information.

## **Conclusions**

Adolescent risk assessment research consistently shows the importance of attending to dynamic changes in risk. Remarkably, however, the present study is one of the few studies to examine this issue. Based on the results, adolescents' risk is indeed dynamic. Contrary to expectations, however, high improvement was generally not associated with lower rates of reoffending. Although this could suggest that the tools are not capturing all relevant changes, a number of equally plausible reasons exist including the fact that change cannot be assumed to be a static entity (i.e., adolescents who show decreases in risk factors during treatment may not necessarily maintain these improvements indefinitely). Further research is needed to clarify the potential value of risk assessment tools in measuring change. Studies that prospectively assess adolescents during and following treatment would be of particular benefit.

### References

- Adam Walsh Child Protection and Safety Act of 2006, 42 U.S.C. § 16901 et seq. (2006).
- Aebi, M., Plattner, B., Steinhausen, H. C., & Bessler, C. (2011). Predicting sexual and nonsexual recidivism in a consecutive sample of juveniles convicted of sexual Offences. *Sexual Abuse: A Journal of Research and Treatment*, 23, 456–473. doi:10.1177/1079063210384634
- Andrews, D. A., & Bonta, J. (1995). *Level of Service Inventory-Revised*. Toronto, ON, Canada: Multi-Health Systems.
- Barnett, G. D., Wakeling, H. C., Mandeville-Norden, R., & Rakestrow, J. (2012). How useful are psychometric scores in predicting recidivism for treated sex offenders? *International Journal of Offender Therapy and Comparative Criminology*, 56, 420-446. doi:10.1177/0306624X11403125
- Barnett, G. D., Wakeling, H., Mandeville-Norden, R., & Rakestrow, J. (2013). Does change in psychometric test scores tell us anything about risk of reconviction in sexual offenders? *Psychology, Crime and Law*, 19, 85-110. doi:10.1080/1068316X.2011.607820
- Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology*, 79, 182–192. doi:10.1037/a0022900
- Borum, R., Bartel, P., & Forth, A. (2006). *Manual for the Structured Assessment for Violence Risk in Youth (SAVRY)*. Odessa, FL: Psychological Assessment Resources.
- Bowen, E., Gilchrist, E., & Beech, A. R. (2008). Change in treatment has no relationship with subsequent re-offending in U.K. domestic violence sample: A preliminary study. *International Journal of Offender Therapy and Comparative Criminology*, 52, 598-614. doi:10.1177/0306624X08319419
- Burns, B.J., Hoagwood, K., & Mrazek, P. (1999). Effective treatment for mental disorders in children and adolescents. *Clinical Child and Family Psychology Review*, 2, 199-254. doi: 10.1023/A:1021826216025
- Butler, S., Baruch, G., Hickey, N., & Fonagy, P. (2011). A randomized controlled trial of Multisystemic therapy and a statutory therapeutic intervention for young offenders. *Journal of the American Academy of Child & Adolescent Psychiatry*, 50, 1220–1235.e2. doi:10.1016/j.jaac.2011.09.017
- Caldwell, M. (2010). Study characteristics and recidivism base rates in juvenile sex offender recidivism. *International Journal of Offender Therapy and Comparative Criminology*, 54, 197-212. doi: 10.1177/0306624x08330016
- Caldwell, M. F., Ziemke, M. H., & Vitacco, M. J. (2008). An examination of the Sex Offender Registration and Notification Act as applied to juveniles: Evaluating the ability to predict sexual recidivism. *Psychology, Public Policy, and Law*, 14, 89–114. doi:10.1037/a0013241
- Caldwell, M., Skeem, J., Salekin, R., & Van Rybroek, G. (2006). Treatment response of adolescent offenders with psychopathy features: A 2-year follow-up. *Criminal Justice and Behavior*, 33, 571-596. doi:10.1177/0093854806288176

- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137. Retrieved from <http://psycnet.apa.org/psycinfo/1982-00095-001>
- Cohen, J. (1988). *Statistical power for the behavioral sciences (2<sup>nd</sup> edition)*. Hillsdale, NJ: Erlbaum.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*, 98-104. doi: 10.1037//0021-9010.78.1.98
- DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated Receiver Operating Characteristic curves: A nonparametric approach. *Biometrics, 44*, 837-845. doi:10.2307/2531595
- Douglas, K. S., Ogloff, J. P., Nicholls, T. L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal Of Consulting And Clinical Psychology, 67*(6), 917-930. doi:10.1037/0022-006X.67.6.917
- Draycott, S., Kirkpatrick, T., & Askari, R. (2012). An idiographic examination of patient progress in the treatment of dangerous and severe personality disorder: A reliable change index approach. *Journal of Forensic Psychiatry and Psychology, 23*, 108-124. doi:10.1080/14789949.2010.481720
- Duff, K. (2012). Evidence-based indicators of neuropsychological change in the individual patient: Relevant concepts and methods. *Archives of Clinical Neuropsychology, 27*, 248-261. doi:10.1093/arclin/acr120
- Edens, J. F., Campbell, J. S., & Weir, J. M. (2007). Youth psychopathy and criminal recidivism: A meta-analysis of the psychopathy checklist measures. *Law and Human Behavior, 31*, 53-75. doi:10.1007/s10979-006-9019-y
- Elkovitch, N., Viljoen, J. L., Scalora, M. J., & Ullman, D. (2008). Research report: Assessing risk of reoffending in adolescents who have committed a sexual offense: The accuracy of clinical judgments after completion of risk assessment instruments. *Behavioral Sciences & The Law, 26*(4), 511-528. doi:10.1002/bsl.832
- Evans, C., Margison, F. & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence Based Mental Health, 1*, 70-72. doi: 10.1136/ebmh.1.3.70
- Falkenbach, D. M., Poythress, N. G., & Heide, K. M. (2003). Psychopathic features in a juvenile diversion population: reliability and predictive validity of two self-report measures. *Behavioral Sciences and the Law, 21*, 787-805. doi:10.1002/bsl.562
- Fleming, W. M., Jory, B., & Burton, D. L. (2002). Characteristics of juvenile offenders admitting to sexual activity with nonhuman animals. *Society and Animals, 10*, 31-45. doi: 10.1163/156853002760030860
- Forth, A. E., Kosson, D. S., Hare, R. D. (2003). *The Psychopathy Checklist: Youth Version*. Toronto, ON: Multi-Health Systems.
- Goodman-Delahunty, J. & O'Brien, K. (in press). Parental sexual offending: Managing risk

- through diversion. *Trends and Issues in Crime and Criminal Justice*. Retrieved from <http://search.informit.com.au/documentSummary;dn=636884774211141;res=IELHSS>> EI SSN: 1836-2206.
- Goeman, J., Meijer, R., & Chaturvedi, N. (2015). Package 'penalized.' Retrieved from <http://cran.r-project.org/web/packages/penalized/penalized.pdf>
- Gretton, H. M., McBride, M., Hare, R. D., O'Shaughnessy, R., & Kumka, G. (2001). Psychopathy and recidivism in adolescent sex offenders. *Criminal Justice and Behavior*, 28, 427-449. doi: 10.1177/009385480102800403
- Guy, L. S. (2008). *Performance indicators of the structured professional judgment approach for assessing risk for violence to others: A meta-analytic survey*. Unpublished doctoral dissertation, Department of Psychology, Simon Fraser University, Burnaby, British Columbia, Canada.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, 143, 29-36. doi:10.1148/radiology.143.1.7063747
- Hare, R. (2003). *Manual for the Hare Psychopathy Checklist*. Toronto, ON: Multi-Health Systems.
- Hare, R. D. (1991). *Manual for the Revised Psychopathy Checklist*. Toronto, ON: Multi-Health Systems.
- Heinze, G. (2006). A comparative investigation of methods for logistic regression with separated or nearly separated data. *Statistics in Medicine*, 25, 4216-4226. doi:10.1002/sim.2687
- Heinze, G., & Ploner, M. (2004). Technical Report 2/2004: A SAS-macro, S-PLUS library and R package to perform logistic regression without convergence problems. Section of Clinical Biometrics, Department of Medical Computer Sciences, Medical University of Vienna, Vienna, Austria.
- Hempel, I., Buck, N., Cima, M., & van Marle, H. (2013). Review of risk assessment instruments for juvenile sex offenders: What is next? *International Journal of Offender Therapy and Comparative Criminology*, 57, 208-228. doi:10.1177/0306624X11428315
- Hilterman, E. (2014, June). *Measuring change over time: Latent growth trajectories with SAVRY risk and protective factors*. Paper presented at the Annual Conference of the International Association of Forensic Mental Health Services, Toronto, Canada.
- Jacobson, N. S., & Truax, P. (1991). Clinical significance : A statistical approach to defining meaningful change in psychotherapy research. *Methodological Issues and Strategies in Clinical Research*, 59, 631-648. doi:10.1037/10109-042
- King, G., & Zeng, L. (2001). Logistic regression in rare events data. *Political Analysis*, 9, 137-163. doi: 10.1093/oxfordjournals.pan.a004868
- Kroner, D. G., & Yessine, A. K. (2013). Changing risk factors that impact recidivism: In search of mechanisms of change. *Law and Human Behavior*, 37, 321-336. doi:10.1037/lhb0000022
- Labrecque, R. M., Smith, P., Lovins, B. K., & Latessa, E. J. (2014). The importance of reassessment: How changes in the LSI-R risk score can improve the prediction of

- recidivism. *Journal of Offender Rehabilitation*, 53, 116-128.  
doi:10.1080/10509674.2013.868389
- Lambert, M. J., Burlingame, G. M., Umphress, V., Hansen, N. B., Vermeersch, D. A., Clouse, G. C., & Yanchar, S. C. (1996). The reliability and validity of the Outcome Questionnaire. *Clinical Psychology and Psychotherapy*, 3, 249-258. doi: 10.1002/(sici)1099-0879(199612)3:4<249::aid-cpp106>3.3.co;2-j
- Latzman, N. E., Viljoen, J. L., Scalora, M. J., & Ullman, D. (2011). Sexual offending in adolescence: A comparison of sibling offenders and nonsibling offenders across domains of risk and treatment need. *Journal Of Child Sexual Abuse: Research, Treatment, & Program Innovations For Victims, Survivors, & Offenders*, 20(3), 245-263.  
doi:10.1080/10538712.2011.571233
- Letourneau, E. J., Armstrong, K. S., Bandyopadhyay, D., & Sinha, D. (2013). Sex offender registration and notification policy increases juvenile plea bargains. *Sexual Abuse: Journal Of Research And Treatment*, 25(2), 189-207. doi:10.1177/1079063212455667
- Letourneau, E. J., Henggeler, S. W., McCart, M. R., Borduin, C. M., Schewe, P. A., & Armstrong, K. S. (2013). Two-year follow-up of a randomized effectiveness trial evaluating MST for juveniles who sexually offend. *Journal of Family Psychology*, 27, 978-985. doi:10.1037/a0034710
- Lewis, K., Olver, M. E., & Wong, S. P. (2013). The Violence Risk Scale: Predictive validity and linking changes in risk with violent recidivism in a sample of high-risk offenders with psychopathic traits. *Assessment*, 20, 150-164. doi:10.1177/1073191112441242
- Lodewijks, H. B., de Ruiter, C., & Doreleijers, T. H. (2010). The impact of protective factors in desistance from violent reoffending: A study in three samples of adolescent offenders. *Journal of Interpersonal Violence*, 25, 568-587. doi:10.1177/0886260509334403
- Lodewijks, H. P. B., Doreleijers, T. A. H., de Ruiter, C., & Borum, R. (2008). Predictive validity of the Structured Assessment of Violence Risk in Youth (SAVRY) during residential treatment. *International Journal of Law and Psychiatry*, 31, 263-271.  
doi:10.1016/j.ijlp.2008.04.009
- Loeber, R., Farrington, D. P., Stouthamer-Loeber, M., & White, H. R. (2008). *Violence and serious theft: Development and prediction from childhood to adulthood*. New York, NY: Routledge.
- Manders, W. A., Deković, M., Asscher, J. J., van der Laan, P. H., & Prins, P. J. M. (2013). Psychopathy as predictor and moderator of Multisystemic therapy outcomes among adolescents treated for antisocial behavior. *Journal of Abnormal Child Psychology*, 41, 1121-1132. doi:10.1007/s10802-013-9749-5
- Marsden, J., Eastwood, B., Wright, C., Bradbury, C., Knight, J., & Hammond, P. (2011). How best to measure change in evaluations of treatment for substance use disorder. *Addiction*, 106, 294-302. doi:10.1111/j.1360-0443.2010.03143.x
- Marshall, J., Egan, V., English, M., & Jones, R. M. (2006). The relative validity of psychopathy versus risk/needs-based assessments in the prediction of adolescent offending behaviour. *Legal and Criminological Psychology*, 11, 197-210.  
doi:10.1348/135532505X68719

- Martinez, R., Flores, J., & Rosenfeld, B. (2007). Validity of the Juvenile Sex Offender Assessment Protocol II (J-Soap-II) in a sample of urban minority youth. *Criminal Justice and Behavior, 34*, 1284–1295. doi:10.1177/0093854807301791
- McGrath, R., J., Cumming, G. F., Burchard, B. L., 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.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*, 30–46. doi:10.1037/1082-989x.1.1.30
- Moffitt, T. E. (1993). Adolescence-limited and life-course-persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*, 674–701. doi:10.1037/0033-295x.100.4.674
- Monahan, K. C., Steinberg, L., Cauffman, E., & Mulvey, E. P. (2013). Psychosocial (im)maturity from adolescence to early adulthood: Distinguishing between adolescence-limited and persisting antisocial behavior. *Development and Psychopathology, 25*, 1093-1105. doi:10.1017/S0954579413000394
- Nickerson, A. B., Colby, S. A., Brooks, J. L., Rickert, J. M., & Salamone, F. J. (2007). Transitioning youth from residential treatment to the community: A preliminary investigation. *Child & Youth Care Forum, 36*(2-3), 73-86. doi:10.1007/s10566-007-9032-4
- O’Neill, M. L., Lidz, V., & Heilbrun, K. (2003). Adolescents with psychopathic characteristics in a substance abusing cohort: Treatment process and outcomes. *Law and Human Behavior, 27*, 299–313. doi:10.1023/a:1023435924569
- Olver, M. E., Lewis, K., & Wong, S. P. (2013). Risk reduction treatment of high-risk psychopathic offenders: The relationship of psychopathy and treatment change to violent recidivism. *Personality Disorders: Theory, Research, and Treatment, 4*, 160-167. doi:10.1037/a0029769
- Olver, M. E., Nicholaichuk, T. P., Kingston, D. A., & Wong, S. P. (2014). A multisite examination of sexual violence risk and therapeutic change. *Journal of Consulting and Clinical Psychology, 82*, 312-324. doi:10.1037/a0035340
- Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2009). Risk assessment with young offenders a meta-analysis of three assessment measures. *Criminal Justice and Behavior, 36*, 329-353. doi: 10.1177/0093854809331457
- Olver, M., Wong, S., 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. doi: 10.1037/1040-3590.19.3.318
- Prentky, R., & Righthand, S. (2003). *Juvenile Sex Offender Assessment Protocol II (J-SOAP-II) Manual*. Washington D.C.: U.S. Department of Justice, Office of Justice Programs, Office of Juvenile Justice and Delinquency Prevention.
- R Core Team (2014). *R: A language and environment for statistical computing reference index*. Retrieved from [http://web.mit.edu/r\\_v3.0.1/fullrefman.pdf](http://web.mit.edu/r_v3.0.1/fullrefman.pdf)
- Rajlic, G., & Gretton, H. M. (2010). An examination of two sexual recidivism risk measures in

- adolescent offenders: The moderating effect of offender type. *Criminal Justice and Behavior*, 37, 1066–1085. doi:10.1177/0093854810376354
- Reh fuss, M. C., Underwood, L. A., Enright, M., Hill, S., Marshall, R., Tipton, P., West, K., & Warren, K. (2013). Treatment impact of an integrated sex offender program as measured by J-SOAP-II. *Journal of Correctional Health Care*, 19, 113-123. doi:10.1177/1078345812474641
- Reitzel, L. R., & Carbonell, J. L. (2006). The effectiveness of sexual offender treatment for juveniles as measured by recidivism: A meta-analysis. *Sexual Abuse: Journal of Research and Treatment*, 18, 401-421. doi:10.1177/107906320601800407
- Rice, M. E., & Harris, G. T. (2005). Violent recidivism: Assessing predictive validity: Correction to Rice and Harris (1995). *Journal of Consulting and Clinical Psychology*, 73, 762–762. doi:10.1037/0022-006x.73.4.762
- Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J., & Müller, M. (2011). pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, 1277. doi:10.1186/1471-2105-12-77
- Salekin, R. T., Leistico, A. M. R., Neumann, C. S., DiCicco, T. M., & Duros, R. L. (2004). Psychopathy and comorbidity in a young offender sample: Taking a closer look at psychopathy's potential importance over disruptive behavior disorders. *Journal of Abnormal Psychology*, 113, 416-427. doi: 10.1037/0021-843x.113.3.416
- Salekin, R. T., Tippey, J. G., & Allen, A. D. (2012). Treatment of conduct problem youth with interpersonal callous traits using mental models: Measurement of risk and change. *Behavioral Sciences and the Law*, 30, 470-486. doi:10.1002/bsl.2025
- Salekin, R. T., Worley, C., & Grimes, R. D. (2010). Treatment of psychopathy: A review and brief introduction to the mental model approach for psychopathy. *Behavioral Sciences and the Law*, 28, 235-266. doi:10.1002/bsl.928
- Serin, R. C., Gobeil, R., & Preston, D. L. (2009). Evaluation of the persistently violent offender treatment program. *International Journal of Offender Therapy and Comparative Criminology*, 53, 57-73. doi:10.1177/0306624X07313985
- Seto, M. C., & Lalumière, M. L. (2010). What is so special about male adolescent sexual offending? A review and test of explanations through meta-analysis. *Psychological Bulletin*, 136, 526-575. doi:10.1037/a0019700
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420–428. doi:10.1037/0033-2909.86.2.420
- Singh, J. P., Grann, M., & Fazel, S. (2011). A comparative study of violence risk assessment tools: A systematic review and metaregression analysis of 68 studies involving 25,980 participants. *Clinical Psychology Review*, 31, 499–513. doi:10.1016/j.cpr.2010.11.009
- Singh, J. P., Yang, S., Mulvey, E. P., & The RAGEE Group (in press). Reporting Guidance for Violence Risk Assessment Predictive Validity Studies: The RAGEE Statement. *Law and Human Behavior*. doi.org/10.1037/lhb0000090
- Skeem, J. L., Polaschek, D. L., Patrick, C. J., & Lilienfeld, S. O. (2011). Psychopathic personality: Bridging the gap between scientific evidence and public

- policy. *Psychological Science in the Public Interest*, 12, 95-162.  
doi:10.1177/1529100611426706
- Spain, S. E., Douglas, K. S., Poythress, N. G., & Epstein, M. (2004). The relationship between psychopathic features, violence and treatment outcome: The comparison of three youth measures of psychopathic features. *Behavioral Sciences and the Law*, 22, 85-102. doi: 10.1002/bsl.576
- Spice, A., Viljoen, J. L., Lutzman, N. E., Scalora, M. J., & Ullman, D. (2013). Risk and protective factors for recidivism among juveniles who have offended sexually. *Sexual Abuse: Journal Of Research And Treatment*, 25(4), 347-369.
- Stein, J., Luppá, M., Brähler, E., König, H., & Riedel-Heller, S. G. (2010). The assessment of changes in cognitive functioning: Reliable change indices for neuropsychological instruments in the elderly—A systematic review. *Dementia and Geriatric Cognitive Disorders*, 29, 275-286. doi:10.1159/000289779
- Strauss, E., Sherman, E. S., & Spreen, O. (2006). *A compendium of neuropsychological tests: Administration, norms, and commentary (3rd edition)*. New York, NY: Oxford University Press.
- Sweeten, G., Piquero, A. R., & Steinberg, L. (2013). Age and the explanation of crime, revisited. *Journal of Youth and Adolescence*, 42, 921-938. doi:10.1007/s10964-013-9926-4
- Viljoen, J. L., Beneteau, J. L., Gulbransen, E., Brodersen, E., Desmarais, S. L., Nicholls, T. L., & Cruise, K. R. (2012a). Assessment of multiple risk outcomes, strengths, and change with the START:AV: A short-term prospective study with adolescent offenders. *International Journal of Forensic Mental Health*, 11, 165–180. doi:10.1080/14999013.2012.737407
- Viljoen, J. L., McLachlan, K., & Vincent, G. M. (2010). Assessing violence risk and psychopathy in juvenile and adult offenders: A survey of clinical practices. *Assessment*, 17, 377-395. doi:10.1177/1073191109359587
- Viljoen, J. L., Mordell, S., & Beneteau, J. L. (2012b). Prediction of adolescent sexual reoffending: A meta-analysis of the J-SOAP-II, ERASOR, J-SORRAT-II, and Static-99. *Law and Human Behavior*, 36, 423–438. doi:10.1037/h0093938
- Viljoen, J. L., Elkovitch, N., Scalora, M. J., & Ullman, D. (2009). Assessment of reoffense risk in adolescents who have committed sexual offenses: Predictive validity of the ERASOR, PCL:YV, YLS/CMI, and Static-99. *Criminal Justice And Behavior*, 36(10), 981-1000. doi:10.1177/0093854809340991
- Viljoen, J. L., Scalora, M., Cuadra, L., Bader, S., Chavez, V., Ullman, D., & Lawrence, L. (2008). Assessing risk for violence in adolescents who have sexually offended: A Comparison of the J-SOAP-II, J-SORRAT-II, and SAVRY. *Criminal Justice and Behavior*, 35, 5–23. doi:10.1177/0093854807307521
- Viljoen, J. L., Nicholls, T., Cruise, K., Desmarais, S., & Webster, C. D. (2014). *Short-Term Assessment of Risk and Treatability: Adolescent Version, User Guide*. Burnaby, BC: Mental Health and Law Policy Institute, Simon Fraser University.
- Vincent, G. M., Guy, L. S., Gershenson, B. G., & McCabe, P. (2012). Does risk assessment make a difference? Results of implementing the SAVRY in juvenile probation. *Behavioral Sciences and the Law*, 30, 384–405. doi:10.1002/bsl.2014

- Vincent, G. M., Guy, L. S., Grisso, T. (2012). *Risk assessment in juvenile justice: A guidebook for implementation*. Chicago, IL: John D. and Catherine T. MacArthur Foundation.
- Vincent, G. M., Terry, A. M., & Maney, S. M. (2009). Risk/needs tools for antisocial behavior and violence among youthful populations. In J. T. Andrade (Ed.), *Handbook of violence risk assessment and treatment: New approaches for mental health professionals* (pp. 337-424). New York, NY: Springer.
- Vose, B., Lowenkamp, C. T., Smith, P., & Cullen, F. T. (2009). Gender and the predictive validity of the LSI-R: A study of parolees and probationers. *Journal of Contemporary Criminal Justice*, 25, 459-471. doi:10.1177/1043986209344797
- Vose, B., Smith, P., & Cullen, F. T. (2013). Predictive validity and the impact of change in total LSI-R score on recidivism. *Criminal Justice and Behavior*, 40, 1383-1396. doi: 10.1177/0093854813508916
- Ward, T. (2002). Good lives and the rehabilitation of sexual offenders: Promises and problems. *Aggression and Violent Behavior*, 7, 513-528. doi: 10.4135/9781483328751.n2
- Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. *Psychology, Crime and Law*, 10, 243-257. doi: 10.1080/1068316041000166274
- White, S. F., Frick, P. J., Lawing, K., & Bauer, D. (2012). Callous-unemotional traits and response to Functional Family Therapy in adolescent offenders. *Behavioral Sciences and the Law*, 31, 271-285. doi:10.1002/bsl.2041
- Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment*, 82, 50-59. doi:10.1207/s15327752jpa8201\_10
- Woessner, G., & Schwedler, A. (2014). Correctional treatment of sexual and violent offenders: Therapeutic change, prison climate, and recidivism. *Criminal Justice and Behavior*, 41, 862-879. doi:10.1177/0093854813520544
- Wong, S., Lewis, K., Stockdale, K., & Gordon, A. (2011). *The Violence Risk Scale-Youth Version*. Unpublished manuscript. University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- Wong, S., & Gordon, A. E. (2003). *The Violence Risk Scale*. Unpublished manuscript, University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- Wong, S., & Hare, R. D. (2005). *Program guidelines for the institutional treatment of violent psychopaths*. Toronto, ON: Multi-Health Systems.
- Wong, S., Olver, M. E., Nicholaichuk, T. P., & Gordon, A. (2003). *The Violence Risk Scale: Sexual Offender Version (VRS: SO)*. Saskatoon, SK: Regional Psychiatric Centre and University of Saskatchewan.
- Worling, J. R., & Curwen, T. (2001). *Estimate of Risk of Adolescent Sexual Offense Recidivism, Version 2.0*. Toronto, Canada: Ontario Ministry of Community and Social Services.
- Worling, J. R., Litteljohn, A., & Bookalam, D. (2010). 20-year prospective follow-up study of specialized treatment for adolescents who offended sexually. *Behavioral Sciences and the Law*, 28, 46-57. doi:10.1002/bsl.912

Table 1

*Interrater Reliability and Internal Consistency of J-SOAP-II and SAVRY Scores*

Measure	Internal Consistency (Cronbach's alpha)		Interrater Reliability ( <i>n</i> = 37)			Increase or Decrease Required to Classify Change as Reliable (95% CI)	
	Admission	Discharge	Admission	Discharge	Change	RCI <sub>α</sub>	RCI <sub>IRR</sub>
<b>J-SOAP-II</b>							
Intervention	.77	.89	.64	.82	.82	4	5
Community Stability/ Adjustment	.60	.72	.52	.77	.64	4	4
Dynamic Risk Total	.80	.91	.65	.84	.82	6	8
<b>SAVRY</b>							
Social/Contextual	.62	.68	.70	.89	.46	4	4
Individual/Clinical	.73	.85	.73	.86	.71	5	5
Protective Factors	.58	.68	.68	.62	.24	3	2
Dynamic Risk Total	.80	.87	.70	.88	.66	7	8

*Note.* The column titled *Change* refers to the interrater reliability of the change scores (i.e., Change Score = Score at Admission minus Score at Discharge for risk scales and Score at Discharge minus Score at Admission for the SAVRY protective factors section). RCI<sub>α</sub> = reliable change index based on internal consistency; RCI<sub>IRR</sub> = reliable change index based on interrater reliability coefficient. RCIs are rounded up or down to the nearest integer, as scale scores are in full numbers versus decimal points.

Table 2

*Stability and Change in J-SOAP-II and SAVRY Scores*

Measure	Admission	Discharge	$F(1, 160)$	$d_{rm}$	$r$
	$M (SD)$	$M (SD)$			
<b>J-SOAP-II</b>					
Intervention	11.26 (2.43)	7.64 (3.63)	260.43***	1.11	.62***
Community Stability/ Adjustment	6.57 (1.91)	4.53 (2.22)	234.14***	0.98	.67***
Dynamic Risk Total	17.83 (3.85)	12.17 (5.56)	-	1.12	.66***
<b>SAVRY</b>					
Social/Contextual	7.04 (2.37)	5.39 (2.37)	148.80***	0.69	.74***
Individual/Clinical	10.67 (3.13)	7.22 (3.81)	253.21***	0.97	.71***
Protective Factors	1.29 (1.29)	1.94 (1.68)	55.16***	0.41	.75***
Dynamic Risk Total	17.71 (4.94)	12.61 (5.70)	-	0.94	.72***

*Note.* \*\*\*  $p < .001$  (two-tailed test).  $d_{rm}$  = repeated-measures Cohen's  $d$ ;  $r$  = stability coefficient.  $F$ -tests were adjusted for their respective family-wise error rate and are significant at  $p < .001$  level. Dynamic Risk Total Scores were not included in the MANOVA due to multicollinearity with the scales.

Table 3

*Proportion of Youth Showing Reliable Change in J-SOAP-II and SAVRY Scores*

Measure	Reliable Decrease <i>n</i> (%)		Reliable Increase <i>n</i> (%)		No Reliable Change <i>n</i> (%)	
	RCI <sub>α</sub>	RCI <sub>IRR</sub>	RCI <sub>α</sub>	RCI <sub>IRR</sub>	RCI <sub>α</sub>	RCI <sub>IRR</sub>
<b>J-SOAP-II</b>						
Intervention	82 (50.3)	64 (39.3)	0 (0.0)	0 (0.0)	81 (49.7)	99 (60.7)
Community Stability/Adjustment	31 (19.0)	31 (19.0)	0 (0.0)	0 (0.0)	132 (81.0)	132 (81.0)
Dynamic Risk Total	94 (57.7)	72 (44.2)	0 (0.0)	0 (0.0)	69 (42.3)	91 (55.8)
<b>SAVRY</b>						
Social/Contextual	11 (6.7)	24 (14.7)	0 (0.0)	0 (0.0)	152 (93.3)	139 (85.3)
Individual/Clinical	63 (38.7)	63 (38.7)	0 (0.0)	0 (0.0)	100 (61.3)	100 (61.3)
Protective Factors	0 (0.0)	0 (0.0)	13 (8.0)	13 (8.0)	150 (92.0)	150 (92.0)
Dynamic Risk Total	61 (37.4)	51 (31.3)	0 (0.0)	0 (0.0)	102 (62.6)	112 (68.7)

*Note.* RCI<sub>α</sub> = reliable change index based on internal consistency; RCI<sub>IRR</sub> = reliable change index based on interrater reliability coefficient.

Table 4

*Relationship between J-SOAP-II Scores and Reoffending: Point-Biserial Correlations and AUCs*

J-SOAP-II	Sexual Reoffense			Violent Nonsexual Reoffense			Any Reoffense		
	<i>r<sub>pb</sub></i>	AUC	95% CI	<i>r<sub>pb</sub></i>	AUC	95% CI	<i>r<sub>pb</sub></i>	AUC	95% CI
<b>Intervention</b>									
Admission	-.01	.46	[.28, .65]	.14	.62	[.49, .74]	.10	.54	[.45, .62]
Discharge	.03	.54	[.39, .69]	.02	.50	[.35, .65]	.03	.51	[.42, .60]
	<i>z</i> = -0.94, <i>p</i> > .05			<i>z</i> = 1.56, <i>p</i> > .05			<i>z</i> = 0.57, <i>p</i> > .05		
<b>Community Stability/Adjustment</b>									
Admission	.00	.50	[.32, .68]	.15*	.63	[.49, .76]	.21*	.61*	[.52, .69]
Discharge	-.03	.46	[.28, .64]	.12	.58	[.43, .72]	.11	.56	[.47, .65]
	<i>z</i> = 0.37, <i>p</i> > .05			<i>z</i> = 0.62, <i>p</i> > .05			<i>z</i> = 1.27, <i>p</i> > .05		
<b>Dynamic Risk Total</b>									
Admission	-.01	.47	[.29, .66]	.16*	.64*	[.52, .77]	.17*	.58	[.49, .67]
Discharge	.01	.51	[.35, .67]	.06	.53	[.37, .68]	.07	.53	[.44, .62]
	<i>z</i> = -0.37, <i>p</i> > .05			<i>z</i> = 1.41, <i>p</i> > .05			<i>z</i> = 1.28, <i>p</i> > .05		

*Note.* \* *p* < .05 (two-tailed test).

Table 5

*Relationship between SAVRY Scores and Reoffending: Point-Biserial Correlations and AUCs*

SAVRY	Sexual Reoffense			Violent Nonsexual Reoffense			Any Reoffense		
	<i>r<sub>pb</sub></i>	AUC	95% CI	<i>r<sub>pb</sub></i>	AUC	95% CI	<i>r<sub>pb</sub></i>	AUC	95% CI
<b>Social/Contextual</b>									
Admission	.04	.56	[.37, .76]	.10	.59	[.45, .73]	.18*	.59*	[.51, .68]
Discharge	.00	.50	[.30, .71]	.12	.57	[.44, .71]	.17*	.59*	[.51, .68]
	<i>z</i> = 0.71, <i>p</i> > .05			<i>z</i> = 0.32, <i>p</i> > .05			<i>z</i> = 0.01, <i>p</i> > .05		
<b>Individual/Clinical</b>									
Admission	-.11	.35	[.20, .50]	.15	.61*	[.50, .72]	.18*	.60*	[.51, .68]
Discharge	-.04	.46	[.31, .61]	.07	.56	[.42, .69]	.09	.55	[.46, .64]
	<i>z</i> = -1.57, <i>p</i> > .05			<i>z</i> = 1.06, <i>p</i> > .05			<i>z</i> = 1.44, <i>p</i> > .05		
<b>Protective Factors<sup>a</sup></b>									
Admission	-.06	.57	[.42, .71]	-.00	.49	[.36, .62]	-.14	.56	[.47, .64]
Discharge	-.05	.53	[.39, .66]	-.01	.52	[.38, .66]	-.11	.55	[.47, .64]
	<i>z</i> = 0.90, <i>p</i> > .05			<i>z</i> = -0.48, <i>p</i> > .05			<i>z</i> = 0.09, <i>p</i> > .05		
<b>Dynamic Risk Total</b>									
Admission	-.06	.44	[.27, .62]	.14	.61	[.48, .73]	.20*	.61*	[.52, .69]
Discharge	-.03	.46	[.28, .64]	.10	.57	[.44, .71]	.13	.57	[.48, .65]
	<i>z</i> = -0.19, <i>p</i> > .05			<i>z</i> = 0.56, <i>p</i> > .05			<i>z</i> = 1.12, <i>p</i> > .05		

*Note.* \* *p* < .05 (two-tailed test).

<sup>a</sup> For ease of interpretation, scores on the Protective Factors were reversed for the AUC analyses such that higher scores represent a deficit in protective factors.

Table 6

*Correlations between Change Scores (i.e., Improvement) and Reoffending*

Change scores	Point-Biserial Correlations			Partial Correlation Controlling for Risk on Static Items		
	Violent			Violent		
	Sexual	Nonsexual	Any	Sexual	Nonsexual	Any
<b>J-SOAP-II</b>						
Intervention	-.05	.10	.05	-.04	.11	.06
Community Stability/ Adjustment	.03	.02	.09	.03	.01	.07
Dynamic Risk Total	-.02	.08	.07	-.02	.08	.07
<b>SAVRY</b>						
Social/Contextual	.04	-.02	.02	.02	-.07	-.04
Individual/ Clinical	-.07	.07	.08	-.09	.04	.05
Protective Factors	.01	-.01	-.01	-.00	-.02	-.02
Dynamic Risk Total	-.03	.04	.06	-.05	-.00	.01
Change scores	Partial Correlation Controlling for Admission scores			Partial Correlation Controlling for Treatment		
	Violent			Violent		
	Sexual	Nonsexual	Any	Sexual	Nonsexual	Any
<b>J-SOAP-II</b>						
Intervention	-.04	.09	.04	-.04	.11	.05
Community Stability/ Adjustment	.04	-.02	.04	.04	.03	.09
Dynamic Risk Total	-.02	.07	.06	-.01	.09	.07
<b>SAVRY</b>						
Social/Contextual	.03	-.06	-.05	.05	-.01	.03
Individual/ Clinical	-.06	.05	.05	-.07	.08	.09
Protective Factors	.00	-.01	-.01	.00	-.02	-.01
Dynamic Risk Total	-.02	.01	.02	-.03	.05	.07

*Note.* All point-biserial and partial correlation coefficients are non-significant. Positive change scores indicate greater improvement. Thus, if youth who showed high improvement were less likely to reoffend, a significant inverse correlation between change and reoffending would be expected.

Table 7

*Penalized Logistic Regression Analyses of the Association between Reliable Change (i.e., Improvements) and Sexual and Violent Nonsexual Reoffending*

Measure	Sexual Reoffense					Violent Nonsexual Reoffense				
	<i>B</i>	<i>SE</i>	Wald	<i>Exp(B)</i>	95% CI	<i>B</i>	<i>SE</i>	Wald	<i>Exp(B)</i>	95% CI
<b>J-SOAP-II</b>										
Step 1										
Static	0.03	0.06	0.28	1.03	[0.92, 1.16]	0.09	0.05	3.45	1.09	[0.10, 1.20]
			$\chi^2(1) = 0.32, p = .571$			$\chi^2(1) = 3.50, p = .061$				
Step 2										
Static	0.00	0.06	0.00	1.00	[0.89, 1.13]	0.08	0.05	3.11	1.09	[0.99, 1.21]
Intervention	-1.98	0.88	6.21*	0.14	[0.02, 0.68]	0.56	0.51	1.19	1.75	[0.63, 4.82]
Community Stability/ Adjustment	1.88	0.82	5.24*	6.58	[1.32, 34.22]	0.16	0.59	0.08	1.18	[0.36, 3.61]
			$\chi^2(3) = 7.96, p = .047$			$\chi^2(3) = 5.40, p = .145$				
<b>SAVRY</b>										
Step 1										
Historical	0.10	0.08	1.65	1.11	[0.95, 1.31]	0.14	0.07	4.70*	1.15	[1.01, 1.31]
			$\chi^2(1) = 1.82, p = .178$			$\chi^2(1) = 4.84, p = .028$				
Step 2										
Historical	0.09	0.08	1.19	1.10	[0.93, 1.30]	0.14	0.07	4.54*	1.15	[1.01, 1.33]
Social/Contextual	1.08	0.84	1.48	2.94	[0.51, 17.30]	-0.73	0.71	1.15	0.48	[0.11, 1.77]
Individual/Clinical	-0.66	0.75	0.76	0.52	[0.09, 2.14]	0.63	0.52	1.48	1.88	[0.67, 5.19]
Protective Factors	-0.99	1.51	0.56	0.37	[0.00, 3.61]	0.25	0.81	0.10	1.26	[0.22, 5.45]
			$\chi^2(4) = 3.84, p = .428$			$\chi^2(4) = 7.03, p = .134$				

Note. \*  $p < .05$  (two-tailed test).

Table 8

*Logistic Regression Analyses of the Association between Reliable Change (i.e., Improvements) and Any Reoffending*

Measure	Any Reoffense				95% CI
	<i>B</i>	<i>SE</i>	Wald	<i>Exp(B)</i>	
<b>J-SOAP-II</b>					
Step 1					
Static	0.08	0.03	7.08**	1.09	[1.02, 1.16]
			$\chi^2(1) = 7.52, p = .006$		
Step 2					
Static	0.07	0.03	4.56*	1.07	[1.01, 1.14]
Intervention	0.01	0.37	0.00	1.01	[0.49, 2.08]
Community Stability/Adjustment	1.12	0.46	5.31*	3.06	[1.18, 7.90]
			$\chi^2(3) = 14.40, p = .002$		
<b>SAVRY</b>					
Step 1					
Historical	0.13	0.04	8.12**	1.13	[1.04, 1.24]
			$\chi^2(1) = 8.62, p = .003$		
Step 2					
Historical	0.10	0.05	4.97*	1.11	[1.01, 1.22]
Social/Contextual	0.11	0.52	0.01	1.12	[0.40, 3.12]
Individual/Clinical	0.63	0.38	2.84	1.88	[0.90, 3.94]
Protective Factors	-0.00	0.63	0.00	1.00	[0.29, 3.40]
			$\chi^2(3) = 12.47, p = .014$		

*Note.* \*  $p < .05$ , \*\*  $p < .01$  (two-tailed test).

Table 9

*Point Biserial Correlations between Change Scores (i.e., Improvement) and PCL:YV Scores*

Change scores	PCL:YV				
	Total	Factor 1	Factor 2	Factor 3	Factor 4
<b>J-SOAP-II</b>					
Intervention	-.10	-.00	-.02	-.13	-.10
Community Stability/Adjustment	-.04	-.03	-.02	-.04	.01
Dynamic Risk Total	-.09	-.01	-.02	-.11	-.07
<b>SAVRY</b>					
Social/Contextual	-.10	-.09	-.04	-.03	-.08
Individual/Clinical	.00	-.01	.01	-.03	.05
Protective Factors	-.06	-.06	.11	-.11	-.07
Dynamic Risk Total	-.04	-.05	-.01	-.03	-.00

*Note.* All zero-order correlation coefficients are non-significant. Positive change scores indicate greater improvement. Thus, if youth with high PCL:YV scores showed less improvement, a significant inverse correlation between change and reoffending would be expected.

Table 10

Comparing Change Scores by PCL:YV Low ( $n = 44$ ), Moderate ( $n = 80$ ), and High ( $n = 39$ ) Groups

Change scores	Mean Change Scores by PCL:YV Groups			$F(2, 160)$	Partial Eta Square $\eta_p^2$	$RCI_{IRR}$ $\chi^2(2)$
	Low $M (SD)$	Moderate $M (SD)$	High $M (SD)$			
<b>J-SOAP-II</b>						
Intervention	4.18 (2.79)	3.76 (2.72)	2.69 (3.07)	3.07	.04	1.60
Community Stability/Adjustment	2.07 (1.45)	2.34 (1.76)	1.41 (1.71)	4.03	.05	0.61
Dynamic Risk Total	6.25 (3.74)	6.10 (4.12)	4.10 (4.56)	-	-	2.44
<b>SAVRY</b>						
Social/Contextual	1.86 (1.92)	1.83 (1.68)	1.04 (1.42)	3.35	.04	3.77
Individual/Clinical	3.59 (2.76)	3.71 (2.60)	2.76 (2.89)	1.70	.02	2.39
Protective Factors	0.77 (1.20)	0.64 (1.07)	0.51 (1.07)	0.57	.01	1.14
Dynamic Risk Total	5.46 (4.40)	5.53 (3.83)	3.80 (4.00)	-	-	2.20

*Note.*  $RCI_{IRR}$  = reliable change index based on interrater reliability coefficient. Dynamic Risk Total Scores were not included in the MANOVA due to multicollinearity with the scales.

### Endnotes

<sup>1</sup> The RCI was calculated using the following formula:  $RCI = [(X_2 - X_1) / S_{diff}]$  where  $X_1$  is the J-SOAP-II/SAVRY dynamic scale score at admission and  $X_2$  is the dynamic scale score at discharge (Jacobson & Truax, 1991).  $S_{diff}$  is the standard error of measurement of the two scores and is calculated as:  $S_{diff} = \sqrt{2(S_E)^2}$ , with  $S_E$  (SEM) calculated as:  $SEM = s_x(\sqrt{1 - r_{xx}})$ , where  $s_x$  is the standard deviation of admission scores and  $r_{xx}$  is the reliability of the tool at admission.

<sup>2</sup> In our follow-up analyses, we conducted Cox regressions with a reduced sample size and obtained similar results to the logistic regressions. We used penalized Cox regression in R for the outcomes of sexual and nonsexual violent reoffending (Goeman, Meijer, & Chaturvedi, 2015). Youth who showed a reliable decrease on the Intervention subscale of the J-SOAP-II showed lower rates of sexually reoffending,  $Exp(B) = 0.11, p < .05$ . Also, youth who showed a reliable decrease on the Community Stability/Adjustment subscale had significantly higher rates of sexual reoffending,  $Exp(B) = 5.90, p < .05$ . Finally, there was a trend wherein youth who showed a reliable decrease on the Community Stability/Adjustment subscale displayed somewhat higher rates of any reoffending,  $Exp(B) = 1.83, p = .06$ . The complete results are available from the authors on request.

<sup>3</sup> For the multivariate analyses, dynamic risk scales for the J-SOAP-II and SAVRY were analyzed separately with the Dynamic Risk Totals removed due to multicollinearity.

<sup>4</sup> All of the youth in the sample had been followed for at least 2 years.