Law and Human Behavior, Vol. 28, No. 4, August 2004 (© 2004)

Reliability of Sexually Violent Predator Civil Commitment Criteria in Florida

Jill S. Levenson[1]

*The present study investigated the reliability of Sexually Violent Predator (SVP) civil commitment criteria under Florida's Jimmy Ryce Act. The purpose of the study was to determine if, independently, 2 evaluators would reach the same conclusions about the same client (*n = *295). According to civil commitment criteria outlined by the United States Supreme Court (*Kansas v. Hendricks*, 1997), SVPs must display a mental abnormality predisposing them to sexual violence and a likelihood of future sexual violence. The interrater reliability of 8* DSM-IV *diagnoses applied by evaluators to determine whether a client has a "mental abnormality that predisposes him to sexual violence" was found to be poor to fair (kappa = .23 to .70). The interrater reliability of risk assessment instruments used to determine "likelihood of reoffense" was good (ICC = .77 to .85). The recommendations made by evaluators regarding whether or not to refer a client for civil commitment demonstrated poor reliability (kappa = .54). Implications for practice and policy are explored.*

KEY WORDS: **A1**

The problem of sexual violence has prompted legislators in 16 states to pass Sexually Violent Predator (SVP) statutes which allow for the civil commitment of dangerous sex offenders following incarceration. The U.S. Supreme Court has twice upheld the constitutionality of SVP civil commitment (*Kansas v. Crane*, 2002; *Kansas v. Hendricks*, 1997) and concluded that incapacitation and treatment are both legitimate aims of SVP statutes. The Supreme Court clarified that selected individuals must have a history of criminal sexual behavior, and must meet two other criteria for SVP commitment: a mental abnormality or personality disorder predisposing the individual to sexual violence, and a likelihood of future sexually violent behavior. This study sought to operationalize these two criteria and explore their reliability using a sample of Florida cases.

Because the term "mental abnormality" is not clinically meaningful, forensic evaluators nationwide have generally interpreted the term to mean "mental

[1]School of Social Work, Florida International University, 11200 S.W. 8th Street, ECS 460, Miami, Florida 33199; e-mail: levenson@fiu.edu.

disorder" and have turned to the *DSM-IV* (American Psychiatric Association [APA], 1994) for diagnostic guidance. Doren (2002) acknowledged that *DSM-IV* definitions sometimes differ from statutory language, but he argued that because the *DSM-IV* is the generally accepted diagnostic manual of the mental healthy community, it serves to communicate the existence of the required syndrome. Others have agreed that a mental disorder predisposing an individual to sexual violence is best reflected by the *DSM-IV* diagnostic category of the Paraphilias (Becker & Murphy, 1998). In some cases, a diagnosis of Antisocial Personality Disorder without a comorbid Axis 1 disorder is sufficient to indicate that a sexually dangerous person has a high degree of psychopathy and a pattern of violating others' rights that primarily consists of sex offending behavior (Doren, 2002).

Florida's civil commitment act was passed in 1998 and named the "Jimmy Ryce Act" in memory of a 9-year-old boy who was abducted, sexually assaulted, and murdered. In order to meet Florida Statute 394 criteria for sexual predator commitment, a convicted sex offender must suffer from a "mental abnormality or personality disorder that makes the person likely to engage in acts of sexual violence." The term "mental abnormality" is further defined in Florida Statute 394 as a "mental condition affecting a person's emotional or volitional capacity which predisposes the person to commit sexually violent offenses" (*Jimmy Ryce Act*, 1998). Florida law does not mandate a *DSM-IV* diagnosis per se, however, evaluators are required by the Florida Department of Children and Families, which is the agency responsible for implementing the statute, to provide a *DSM-IV* diagnosis. Without the requisite mental abnormality, an offender cannot be involuntarily committed. In practice, the *DSM-IV* is used by evaluators in Florida to determine whether offenders meet the first criterion for commitment.

The second criterion set forth by the U.S. Supreme Court for civil commitment requires that a convicted sex offender be likely to commit future sexually violent crimes. This likelihood is usually determined through the use of actuarial risk assessment instruments which estimate the probability of reoffense by referring to the known recidivism rates of individuals with similar characteristics. Such instruments cannot be used to predict with certainty that a given individual will behave in a particular way. They can, however, provide important data with which to inform one's expectations regarding the likelihood of a particular outcome.

In Florida, "likely" means that "the person's propensity to commit acts of sexual violence is of such a degree as to pose a menace to the health and safety of others" (*Jimmy Ryce Act*, 1998). The risk assessment instruments most commonly used by SVP evaluators in Florida are the Static-99 (Hanson & Thornton, 1999), the Rapid Risk Assessment for Sex Offense Recidivism (RRASOR; Hanson, 1997), and the Minnesota Sex Offender Screening Tool—Revised (MnSOST-R; Epperson et al., 1999). Although the Psychopathy Checklist Revised (PCL-R; Hare, 1991) is not an actuarial risk assessment instrument, psychopathy is a risk factor that is correlated with sexual reoffense (Quinsey, Harris, Rice, & Cormier, 1998; Rice & Harris, 1997) and therefore the PCL-R is frequently used by SVP evaluators.

Reliability of DSM Diagnostic Criteria

In general, the DSM manuals have demonstrated disappointing reliability in over 20 years of research, achieving an average kappa of only .64 across seven large data sets (reviewed in Meyer, 2002). At the time the *DSM-IV* was released, no reliability data had been published, and although many believe that reliability is still clearly a problem, little new research has focused on it (Kirk & Kutchins, 1994). It has been argued that the *DSM-III* improved the reliability of psychiatric diagnoses, however, others maintain that diagnostic accuracy has not necessarily improved (reviewed in Reid, Wise, & Sutton, 1992). Structured interview protocols result in slightly improved reliability (.72) of Axis 1 disorders (Meyer, 2002).

Several critics have argued that the *DSM-IV* (APA, 1994) paraphilia criteria might not accurately reflect existing theoretical, empirical, clinical, and descriptive literature, and concerns regarding the validity and reliability of such diagnoses have been noted (Doren, 2002; Marshall, 1997; Marshall, Kennedy, & Yates, 2002; O'Donohue, Regev, & Hagstrom, 2000). Only two true reliability studies of paraphilia diagnoses have been attempted (Marshall, Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002) and these have focused exclusively on Sexual Sadism.

The first noted concern is the 6-month time period required to diagnose a paraphilia. It is well-established that sex offenders often commit many sexual crimes for which they are not caught (Abel et al., 1987; Ahlmeyer et al., 2000; Freeman-Longo, 1985; Hindman & Peters, 2000), but it may be difficult for clinicians to diagnose paraphilias in offenders for whom no prior arrest records exist (O'Donohue et al., 2000). The 6-month time frame creates additional problems, because the length of time is arbitrary and may preclude a diagnosis of a paraphilia in an individual who commits multiple sex offenses within the 6 month window (O'Donohue et al., 2000). Conversely, the criteria would include a sex offender who sexually molested one child on two occasions over a time frame that extended beyond 6 months, but who may not have the primary sexual orientation toward children that clinicians believe characterizes a pedophilic disorder.

O'Donohue et al. (2000) pointed out that the terms "recurrent" and "intense" used in *DSM-IV* criteria A (p. 528) to describe the frequency and quality of paraphilic fantasies and behaviors are vague and subjective. They require diagnosticians to rely on inferences that reduce the validity and reliability of the criterion. To further complicate matters, Marshall (1997) argued, offenders are understandably reluctant to admit having deviant fantasies, urges or behaviors, as evidenced by a survey where only 65% of 205 convicted sex offenders admitted to the crime for which they were found guilty (Wormith, 1983).

Criterion B of the *DSM-IV* diagnosis for a paraphilia holds similar ambiguity with the problematic assumption that the disorder will create distress or impairment (O'Donohue et al., 2000). Although an arrest would certainly interfere with social or occupational functioning, it may be the arrest that creates distress for the individual rather than the behavior itself. Although it is true that many sex offenders feel a great deal of shame, guilt, and conflict over their behavior (Marshall, 1997), the behavior of other abusers is ego-syntonic (Salter, 1995). Most notable, for instance,

is the North American Man-Boy Love Association (NAMBLA), an organization whose members are pedophiles that advocate for laws allowing adults to have sex with children. In a study of 50 members of a pedophilic organization, it was found that 90% of participants reported no desire to reduce or eliminate their sexual interest in children (Bernard, 1975). Thus, distress and impairment are in the eye of the beholder, and many pedophiles will perceive or experience neither (O'Donohue et al., 2000; Marshall, 1997). To address these criticisms, changes were made in
A2 the *DSM-IV* Text Revision (APA, 2000) to clarify criteria for Paraphilias, allowing the requirement for clinical significance to be met if the person has engaged in paraphilic behavior even when subjective distress or impairment does not exist. This change has, to some extent, simplified one diagnostic challenge faced by SVP evaluators.

Another diagnostic problem is that the *DSM-IV* (APA, 1994) specifies that Pedophilia generally involves prepubescent children. However, all states have statutes rendering it unlawful for an adult to engage in sexual activity with a child under age 16, and many offenders being considered for civil commitment have adolescent victims. It could be argued that sexual attraction to a developed, physically mature adolescent does not constitute a deviant sexual preference and is therefore not a mental disorder. On the other hand, one study reported that offenders with victims ages 13–15 have higher recidivism rates than molesters of younger age groups (Epperson et al., 1999). Rather than diagnose these offenders as pedophiles, SVP evaluators typically turn to the diagnosis Paraphilia NOS (Hebephilia). The lack of a specific diagnosis for rapists has also resulted in the use of Paraphilia NOS, nonconsent (Doren, 2002).

Two studies have specifically investigated the reliability and validity of the diagnosis of Sexual Sadism. The records of 59 Canadian prisoners diagnosed with sexual sadism were compared with a group who did not receive the diagnosis (Marshall,
A3 Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002). Results revealed no statistically significant differences between the two groups, except on two variables: beating and torture. Surprisingly, it was the nonsadistic group who displayed higher frequencies of these behaviors. A subsequent study (Marshall, Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002) further confirmed that the adjusted percentage of absolute agreement on the diagnosis of Sexual Sadism made by 15 experts rating 12 cases was about 22%, which was equivalent to a kappa of .14. The authors concluded that the reliability of the diagnosis was poor, and caution was raised about the adequacy of the *DSM-IV* diagnostic criteria (Marshall, Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002).

Reliability of Sex Offender Risk Assessment Instruments

Studies have reported consistently favorable results when investigating the reliability of commonly used risk assessment instruments. Recently, in an Arizona study interrater reliabilities of the RRASOR, Static-99, and MnSOST-R were found to be .90, .90, and .90, respectively (Bartosh, Garby, Lewis, & Gray, 2003). Interrater reliability of the RRASOR, Static-99, and MnSOST-R was tested on Canadian samples by calculating the agreement on total scores between two raters ($n = 30$; Barbaree, Seto, Langton, & Peacock, 2001). Results yielded Pearson correlation coefficients

of .94, .90, and .80, respectively. Interrater reliability of the PCL-R was calculated between total scores using two raters, and yielded an correlation coefficient of .81 (Barbaree et al., 2001).

Another recent study investigated the predictive validity of four instruments among four diverse samples of sex offenders from three different sites (Harris, Rice, Quinsey, Lalumiere, Boer, & Lang, in press). Reliability was assessed by comparing scores generated by two independent codings on 10 subjects. Intraclass correlation coefficients were .95 for the RRASOR, and .87 for the Static-99.

An unpublished interrater reliability study of the MnSOST-R used 27 Florida SVP evaluators to rate 10 cases. The agreement among raters for the total scores of all 10 cases was calculated using the intraclass correlation coefficient (ICC = .85; Bradley & Epperson, 2001). Overall, the risk assessment instruments commonly used in SVP evaluations have demonstrated consistently good reliability.

Purpose of the Study

The purpose of this study was to examine the interrater reliability of sex offender civil commitment criteria. By operationalizing the criteria for civil commitment outlined by the U.S. Supreme Court (*Kansas v. Hendricks*, 1997), three specific research questions were posed: (1) What is the interrater reliability of *DSM-IV* diagnoses used to assess whether an offender has a mental abnormality or personality disorder that predisposes him to sexual violence?; (2) What is the interrater reliability of risk assessment instruments used in assessing if an offender is likely to reoffend?; and (3) What is the interrater reliability of the recommendations for civil commitment that were made by evaluators? The study intends to contribute to the limited empirical literature regarding sexually violent predator statutes by exploring whether independent evaluators reach the same conclusions about sex offender clients who are assessed for civil commitment.

These questions were considered to be important and interesting for several reasons. First, few inquiries into the interrater reliability of *DSM-IV* paraphilia diagnoses have been attempted. This study will add to a very limited literature on the reliability of DSM diagnoses. Second, the study contributes to the growing literature assessing the psychometric properties of actuarial sex offender risk assessment instruments commonly utilized across North America. Often in instrumentation studies, subjects are specially trained and factors are controlled to maximize consistency across raters. This study examines the interrater reliability of the instruments in a real-world setting. Finally, and perhaps most important, this study is the first effort to explore the reliability of sex offender civil commitment selection, a controversial issue for policy makers and practitioners alike.

METHOD

Sample

The sample was drawn from the entire population of 450 male, adult, competent, convicted sex offenders in Florida prisons who received a face-to-face evaluation by

psychologists or psychiatrists for SVP civil commitment between July 1, 2000 and June 30, 2001. Consistent with statutory language, many subjects were examined by more than one forensic evaluator. These were the cases that were chosen for analysis, and included 295 cases. All initial evaluations were assigned to a forensic examiner by a private agency providing contracted services. Therefore, the evaluators "worked for" neither the state nor the defense and were, ostensibly, objective. All evaluators did not utilize all available risk assessment instruments in every case, and therefore the sample size varies for each instrument.

The mean age of the 450 men in the sampling frame was 41 years, and subjects had an average of 11.2 years of education. Nearly half (44.5%) belonged to a racial or ethnic minority group. Only 13.5% of the sampling pool was currently married. About 9.2% had no diagnosis. About 30% had one diagnosis, 32% had two diagnoses, and 21% had three diagnoses. Only 7% had more than three diagnoses. Rapists, who were defined as having victims who were all over the age of 18, comprised 25.6% of the sampling frame. About half (49.3%) were child molesters whose victims were all under age 18. Mixed offenders (22%) had both adult and minor victims. Noncontact offenders (2.7%) included perpetrators of exhibitionism, voyeurism, and computer-related sex crimes. The vast majority had extrafamilial victims (92%). Most had only female victims (71%), 10% had only male victims, and 18% had sexually assaulted victims of both sexes. The average number of sex crime arrests was 2.7, with nonsexual arrests averaging 8.1.

Of the 295 offenders included in the current study, the mean age was 41, and the subjects had, on average, an 11th grade education. About 47% belonged to a minority group, and 14% were currently married. Only 5% had no diagnosis, 21% had one diagnosis, 37% had two diagnoses, and 36% had three or more. Rapists accounted for 23% of the sample, 45% were child molesters, 30% were mixed, and 2% were noncontact offenders. About 97% of the sample had extrafamilial victims, with 67% having female victims only, 11% having male victims, and 21% having victims of both sexes. The mean numbers of sexual and nonsexual arrests for the sample were three and nine, respectively. The sample of subjects who received two evaluations seem to be slightly more likely to have multiple DSM diagnoses, to have a variety of offenses, to have both male and female victims, and to have more arrests.

A total of 25 evaluators, of whom 88% were male, examined the subjects. Evaluators were all licensed psychologists or psychiatrists, possessing a PhD (76%), PsyD (16%), or MD (8%). The amount of experience evaluating and/or treating sex offenders prior to being hired to evaluate sex offenders under Florida's Jimmy Ryce Act was known regarding only 11 evaluators. Of those who submitted a C.V. to this investigator upon request, the mean number of years experience was 5.9, with a range of 0–18 years of experience.

Data Collection Procedure

This research was conducted in accordance with prevailing ethical principles and was approved by the University's Institutional Review Board. Data were collected via review of SVP evaluation reports provided by the Florida Department of Children and Families. Evaluations typically included psychosocial information,

criminal history, diagnosis, and risk assessment. Subjects were interviewed by evaluators at the prisons in which they were incarcerated, prior to their scheduled release dates.

Data collected included total scores on four risk assessment instruments: Static-99, RRASOR, Mn-SOST-R, and PCL-R. *DSM-IV* diagnoses were coded dichotomously (yes/no) and included the diagnoses most commonly considered: Pedophilia, Sexual Sadism, Exhibitionism, Paraphilia NOS, Antisocial Personality Disorder, Personality Disorder NOS, Other Personality Disorder, Substance Disorder, and Other Major Mental Illness. Finally, the evaluator's recommendation for civil commitment (yes/no) was recorded.

Data Analysis Procedures

Interrater reliability was examined on total scores of each instrument using the intraclass correlation coefficient (ICC) for continuous variables (Bartko, 1966). Because each case was rated by both evaluators, an ICC two-way model was used (Shrout & Fleiss, 1979). Because all of the cases in the final study were combined for analysis, a random model was appropriate, as judges' effects will contribute to the variability of the ratings (Shrout & Fleiss, 1979). Each item was assessed for total agreement to determine whether the two judges were interchangeable (Shrout & Fleiss, 1979). The *DSM-IV* diagnoses and civil commitment recommendation were dichotomous variables and were therefore assessed using kappa. Although different interpretations of reliability coefficients exist, when critical decisions are to be made, such as civil commitment, higher standards should be used. For this study, a reliability coefficient below .60 is considered poor, .60 to .74 is considered fair, and .75 to 1.0 is considered good (Bloom, Fischer, & Orme, 1999).

RESULTS

All interrater reliability results are displayed in Table 1. Diagnoses did not fare so well; reliability was poor for most diagnoses, with kappa coefficients ranging from .23 to .70. Pedophilia demonstrated fair reliability, as did "other mental illness," which was defined as a psychotic disorder or major mood disorder. Because it is possible that evaluators can agree that an individual meets criteria for some paraphilia (even if they cannot agree on which one) the four paraphilic diagnoses were collapsed into one category. However, reliability was not much improved by this recoding (kappa = .47).

The risk assessment instruments all revealed good reliability. Analysis of the Static-99 produced an ICC coefficient of .85, as did the Mn-SOST-R. The PCL-R was used by both evaluators in only 69 cases, achieving an ICC coefficient of .84. The RRASOR was used by both evaluators in only 16 cases and the reliability coefficient was .77.

Finally, the interrater reliability of the civil commitment recommendation proved to be poor (.54).

Table 1. Reliability Coefficients

	n	ICC	Kappa
Risk Assessment Instruments			
Static-99	281	.85	
MnSost-R	224	.85	
RRASOR	16	.77	
PCL-R	69	.84	
Diagnoses			
Pedophilia	277		.65
Sexual Sadism	277		.30
Exhibitionism	277		.47
Paraphilia NOS	277		.36
Personality Disorder NOS	276		.23
Other Personality Disorder	276		.30
Substance Disorder	277		.43
Other Mental Illness	276		.70
Any paraphilia	277		.47
Civil Commitment Selection			
Recommendation	295		.54

DISCUSSION

This study sought to determine the reliability of civil commitment criteria under SVP statutes in Florida. Specifically, a question of interest was, "when two evaluators assess the same offender, will they come to the same conclusions?" Although evaluators came to the same conclusion more than half the time, according to Bloom et al. (1999), the reliability of the commitment recommendation was poor. Of the two requisite criteria for commitment, the reliability of *mental abnormality or personality disorder*, as measured by relevant *DSM-IV* diagnoses, also proved to be poor. The reliability of the second criterion, *likelihood of reoffense*, as measured by risk assessment scores, demonstrated good interrater reliability. The interrater reliability of the instruments investigated in this study is consistent (albeit, somewhat lower) with other research (Barbaree et al., 2001; Bartosh et al., 2003; Harris et al., in press).

These results are noteworthy because although likelihood of sexual reoffense (as measured by the actuarial risk assessment instruments) is a necessary component for commitment, it is not sufficient alone to justify commitment. Federal case law and Florida's Jimmy Ryce Act require an SVP to suffer from a mental abnormality that predisposes the individual to future sexual violence, and thus diagnosis is an integral component of the assessment of dangerousness. It is possible for offenders to obtain moderate or high scores on risk assessment instruments by virtue of their extensive criminal history or demographic factors. Therefore it has been argued that an individual should meet the criteria for a paraphilia disorder in order to be recommended for SVP commitment (Becker & Murphy, 1998). Without the requisite mental abnormality demonstrating sexual deviance, it cannot be determined that the subject is predisposed to commit a future sexually violent offense as opposed to a generally violent offense.

It is possible that in this sample, the poor reliability was related to the variability in evaluator experience evaluating and treating sex offenders. This explanation seems doubtful, however, given the consistency of the findings with prior research on the risk assessment instruments (Barbaree et al., 2001; Bartosh et al., 2003; Harris et al., in press) and the DSM diagnoses (Meyer, 2002). In other words, the interrater reliability of the instruments was pretty good despite variable assessor experience, and the *DSM-IV* diagnoses was poor, which is similar to the results of other studies examining DSM reliability in general.

The poor reliability of civil commitment selection appears to be related to the ambiguity of DSM diagnostic criteria. Without consistent agreement on the diagnoses that render an individual predisposed to commit future sexually violent crimes, it is unlikely that evaluators will be able to agree on the commitment recommendation. Some scholars have identified practice considerations and proposed tentative solutions for improving diagnostic reliability of the Paraphilias.

Doren (2002) argued that some rapists, even those with multiple arrests for rape, may not be paraphilic. He suggested that those who meet *DSM-IV* (APA, 1994) criteria for Paraphilia NOS, nonconsent, should demonstrate sexual interest in rape as evidenced by any set of possible multiple criteria such as: ejaculation or other signs of sexual arousal during events that are clearly nonconsensual; repetitive patterns of actions; a pattern in which virtually all of the person's criminal behavior is sexual; a pattern of rape when the victim had already been willing to have consensual sex; a short time period after consequence before raping again; committing rape under circumstances with a high likelihood for detection; and having concomitant cooperative sexual partners. Doren (2002) emphasized the need for diagnosticians to differentiate the paraphilic rapist from what might be called a merely opportunistic rapist. Since none of these criteria are stated or implied in the *DSM-IV*, it is not surprising that, in practice, the diagnosis of Paraphilia NOS, nonconsent, is widely variable.

O'Donohue et al. (2000) outlined suggestions for improvement of Pedophilia by proposing a "Pedophilia response disorder" that specifies a single episode as "acute" and a pattern of behavior as "chronic." Such a definition would be more consistent with that of the clinical treatment community, who generally use the label "pedophile" to refer to men whose primary or only sexual orientation is toward children, or whose offense patterns are predatory. In other words, all pedophiles are child molesters but all child molesters are not necessarily pedophiles.

Sexual Sadism also presents diagnostic challenges, not the least of which is that an evaluator must infer sexual arousal to sadistic acts in cases where clients do not readily admit such arousal. In many cases, differential diagnosis must be based on external sources of information such as sexual history, victim partner interviews, and crime scene data (Marshall, Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002), which is often not available to SVP evaluators. Because misdiagnosis of Sexual Sadism has such important implications, high reliability is demanded of the diagnosis (Marshall, Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002). Marshall et al. (Marshall, Kennedy, & Yates, 2002; Marshall, Kennedy, Yates, & Serran, 2002) proposed that a diagnosis of sadism should be reserved for clients

whose violence includes mutilation, cruelty, torture, and for whom evidence exists of deviant sexual arousal to such violence.

In summary, practitioners are faced with diagnostic criteria that contradict both empirical research and clinical conceptualization. It has been suggested that DSM may not be sufficiently clear and accurate for use in legal settings, especially those in which grave consequences are imposed by a judge or jury (Reid et al., 1992). The reasons are varied, and include: the limitations of one clinical interview to adequately assess severity of symptoms, patient dissimulation and manipulation, and incomplete, inaccurate, or irrelevant review of records. Evaluator bias can also affect diagnostic decisions. Because the DSM is the only diagnostic taxonomy recognized by American courts and mental health professionals, its idiosyncrasies and shortcomings have a direct impact on SVP proceedings.

Little research has been done on the process and practice of SVP selection. Preliminary research has found that actuarial risk assessment scores and paraphilia diagnoses were predictive of civil commitment recommendations (Levenson, 2003b, 2003c), and that selected offenders, as a group, differ significantly on these variables from released offenders (Levenson, 2003a). These findings suggested that SVP commitment in Florida is a valid process; it is successful in selecting high risk offenders who are predisposed to sexual violence. Although validity usually implies reliability, the current study found an unacceptable degree of inconsistency among evaluators, particularly related to diagnostic decisions and civil commitment selection.

This study is limited in its ability to generalize these findings to civil commitment processes and practices in other states. Another limitation was that only total actuarial instrument scores were available, so a more detailed analysis of each item on each measure was not possible. Similarly, it would have been revealing to analyze the *DSM-IV* criteria for each diagnosis to determine if particular criterion were more or less reliable than others.

SVP civil commitment evaluations have serious consequences for both clients and communities. Dangers exist at both ends of the continuum: confining individuals who might not reoffend is as problematic as releasing some who may continue to commit sexually violent crimes. Thus, efforts to improve the reliability of assessment criteria are crucial. Continued research is needed to investigate this question across the United States, and to identify problematic trends and propose solutions. The credibility and success of SVP statutes depend inherently on the ability to accurately and reliably select high-risk violent sexual predators for civil commitment.

REFERENCES

Abel, G. G., Becker, J. V., Cunningham-Rathner, J., Mittleman, M. S., Murphy, M. S., & Rouleou, J. L. (1987). Self-reported crimes of nonincarcerated paraphiliacs. *Journal of Interpersonal Violence*, *2,* 3–25.

Ahlmeyer, S., Heil, P., McKee, B., & English, K. (2000). The impact of polygraphy on admissions of victims and offenses in adult sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, *12*(2), 123–138.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th eds.). Washington, DC: Author.

Barbaree, H. E., Seto, M. C., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, *28*(4), 490–521.

Bartko, J. J. (1966). The intraclass correlation coefficient as a measure of reliability. *Psychological Reports*, *19,* 3–11.

Bartosh, D., Garby, T., Lewis, D., & Gray, S. (2003). Differences in the predictive validity of actuarial risk assessments in relation to sex offender type. *International Journal of Offender Therapy and Comparative Criminology*, *47*(4), 422–438.

Becker, J. V., & Murphy, W. (1998). What we know and do not know about assessing and treating sex offenders. *Psychology, Public Policy and Law*, *4*(1/2), 116–137.

Bernard, F. (1975). An enquiry among a group of pedophiles. *Journal of Sex Research*, *11*(3), 242–255.

Bloom, M., Fischer, J., & Orme, J. G. (1999). *Evaluating practice: Guidelines for the accountable professional* (3rd ed.). Boston: Allyn and Bacon.

Bradley, M. A., & Epperson, D. L. (2001). [Interrater reliability study of the MnSOST-R]. Unpublished raw data.

Doren, D. M. (2002). *Evaluating sex offenders: A manual for civil commitments and beyond*. Thousand Oaks, CA: Sage.

Epperson, D. L., Kaul, J. D., Huot, S. J., Hesselton, D., Alexander, W., & Goldman, R. (1999). *Minnesota sex offender screening tool—revised (MnSost-R): Development performance, and recommended risk level cut scores.*, Retrieved from http://psych-server.iastate.edu/faculty/epperson/MnSOST-R.htm

Freeman-Longo, R. E. (1985). *Incidence of self-reported sex crimes among incarcerated rapists and child molesters*. Unpublished manuscript.

Hanson, R. K. (1997). *The development of a brief actuarial scale for sexual offense recidivism*. Ottawa: Department of the Solicitor General of Canada.

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

Hare, R. D. (1991). *The Hare Psychopathy Checklist—Revised*. Toronto, Canada: Multi-Health Systems.

Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumiere, M. L., Boer, D. P., & Lang, C. (in press). A multi-site comparison of actuarial risk instruments for sex offenders. *Psychological Assessment*. **A4**

Hindman, J., & Peters, J. M. (2000). How polygraphing adult sex offenders can lead us to better understanding the juvenile sex offender. *Federal Probation*. **A5**

Jimmy Ryce Involuntary Civil Commitment for Sexually Violent Predators' Treatment and Care Act, Florida Statute 394.912 (1998).

Kansas v. Crane, 534 U.S. 407 (2002).

Kansas v. Hendricks, 117 S. Ct. 2072 (1997).

Kirk, S. A., & Kutchins, H. (1994). The myth of the reliability of DSM. *The Journal of Mind and Behavior*, *15*(1), 71–86.

Levenson, J. S. (2003a). *Civil commitment of Sexually Violent Predators: A comparison of selected and released offenders*. Manuscript submitted for publication.

Levenson, J. S. (2003b). Factors predicting recommendations for civil commitment of sexually violent predators under Florida's Jimmy Ryce Act. *Dissertation Abstracts International*, *64*(03), UMI no. AAT 3085817.

Levenson, J. S. (2003c). *Factors predicting selection of sexually violent predators for civil commitment under Florida's Jimmy Ryce Act*. Manuscript submitted for publication.

Marshall, W. L. (1997). Pedophilia: Psychopathology and theory. In D. R. Laws & W. O'Donohue (Eds.), *Sexual deviance*. New York: Guilford Press.

Marshall, W. L., Kennedy, P., & Yates, P. (2002). Issues concerning the reliability and validity of the diagnosis of sexual sadism applied in prison settings. *Sexual Abuse: A Journal of Research and Treatment*, *14*(4), 301–311.

Marshall, W. L., Kennedy, P., Yates, P., & Serran, G. A. (2002). Diagnosing sexual sadism in sexual offenders: Reliability across diagnosticians. *International Journal of Offender Therapy and Comparative Criminology*, *46*(6), 668–677.

Meyer, G. J. (2002). Implications of information-gathering methods for a refined taxonomy of psychopathology. In L. E. Beutler & M. L. Malik (Eds.), *Rethinking the DSM: A psychological perspective* (pp. 69–106). Washington, DC: American Psychological Association.

O'Donohue, W., Regev, L. G., & Hagstrom, A. (2000). Problems with the *DSM-IV* diagnosis of Pedophilia. *Sexual Abuse: A Journal of Research and Treatment*, *12*(2), 95–105.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.

Reid, W. H., Wise, M., & Sutton, B. (1992). The use and reliability of psychiatric diagnosis in forensic settings. *Clinical Forensic Psychiatry*, *15*(3), 529–537.

Rice, M. E., & Harris, G. T. (1997). Cross-validation and extension of the Violence Risk Appraisal Guide for child molesters and rapists. *Law and Human Behavior*, *21,* 231–241.

A6

Salter, A. (1995). *Transforming trauma*. Thousand Oaks, CA: Sage.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Wormith, J. S. (1983). A survey of incarcerated sexual offenders. *Canadian Journal of Criminology*, *25*(4), 379–390.

Queries to Author:

A1. Au: Kindly provide atleast 4–6 keywords for this article.
A2. Au: Kindly add "APA, 2000" in the Ref. list.
A3. Au: Kindly note that "Marshall et al., 2002" has been replaced with "all the authors" with year 2002 to differentiate both references. Please check (page nos. 8, 9, 18, and also page no. 6) and delete whichever is not applicable.
A4. Au: Kindly update this ref.
A5. Au: Kindly provide the volume no. and page range for this ref.
A6. Au: Kindly check the page range.