

Sadism among Sexual Homicide Offenders: Validation of the Sexual Sadism Scale

(BRIEF REPORT)

Ewa B. Stefanska

University of Roehampton

Joachim Nitschke

Ansbach District Hospital

Adam J. Carter

HM Prison and Probation Service

Andreas Mokros

University Hospital of Psychiatry Zurich and FernUniversität in Hagen

Author Note

Ewa B. Stefanska, Department of Psychology, University of Roehampton, London, UK.

Joachim Nitschke, Forensic-Psychiatric Hospital, Ansbach District Hospital, Ansbach,

Germany. Adam J. Carter, HM Prison and Probation Service, London, UK. Andreas Mokros;

Department of Forensic Psychiatry, University Hospital of Psychiatry Zurich, Zurich,

Switzerland, and Department of Psychology, FernUniversität in Hagen (University of

Hagen), Hagen, Germany.

Andreas Mokros is now at the FernUniversität in Hagen (University of Hagen), Germany.

The research reported herein was presented at the 2<sup>nd</sup> International Convention of

Psychological Science (ICPS), Vienna, Austria, on March 25, 2017. The authors would like

to thank Professor Dimiter M. Dimitrov, Ph.D., from George Mason University, Fairfax, VA,

for his help.

Correspondence concerning this article should be addressed to Andreas Mokros, Department

of Psychology, FernUniversität in Hagen (University of Hagen), 58084 Hagen, Germany. E-

mail: andreas.mokros@fernuni-hagen.de

Sadism among Sexual Homicide Offenders: Validation of the Sexual Sadism Scale  
(BRIEF REPORT)

### **Abstract**

Sexual sadism is assumed to be a crucial factor in sexual homicide. Prevalence estimates vary greatly due to differences in the definition of sexual sadism. A nationwide sample of 350 male perpetrators who had committed a sexual homicide offense against a female aged 14 years or above in England or Wales was assessed based on archival records. Sexual sadism was assessed using the Sexual Sadism Scale (SeSaS; Mokros et al., 2014, *Psychological Assessment*, 26, 138-147). Item response theory (IRT) analyses were conducted focusing on the two-parameter logistic model. The single-factor structure of the SeSaS Part 1 was tested using confirmatory factor analysis. Estimates of both internal consistency and interrater agreement were satisfactory to substantial. IRT analysis showed that the Part 1 items captured moderate to severe levels of the latent construct (i.e., theta levels  $> 0$ ). Based on the Posterior Probability of Diagnosis (PPOD) index (Lindhiem et al., 2013, *Psychological Assessment*, 25, 456-466) the prevalence of the disorder was estimated at 37% in the sample. The substantial correlation between the SeSaS Part 1 total score and original clinical diagnoses of sadism confirms the criterion validity of the scale. Exertion of control and infliction of torture were among the most informative items. In sum, the results support the usefulness of the SeSaS instrument for assessing forensically relevant forms of sadism.

Public Significance Statement: Sexual sadism is highly prevalent among the perpetrators of sexual homicide. In sexual homicide offenders, a checklist based on crime-scene behavior proves helpful to establish a tentative diagnosis.

*Keywords:* sadism, sexual homicide, prevalence, SeSaS, PPOD

## Sadism among Sexual Homicide Offenders: Validation of the Sexual Sadism Scale

Historically, sadism has been conceptualized as follows:

Sadism is the experience of sexual pleasurable sensations (including orgasm) produced by acts of cruelty, bodily punishment afflicted on one's own person or when witnessed in others, be they animals or human beings. It may also consist of an innate desire to humiliate, hurt, wound or even destroy others in order thereby to create sexual pleasure in one's self. (von Krafft-Ebing, 1906, p. 80).

Since then various definitions have been introduced with criteria that often disagree on the primary motivating force that drives sexual sadists (Marshall & Kennedy, 2003). Such assumptions on motivating forces include: Humiliation of the victim (e.g. Ressler et al., 1988; Warren, Hazelwood & Dietz, 1996); control of the victim (e.g. MacCulloch, Snowden, Wood, & Mills, 1983); the use of aggression (e.g. Myers, Burgess, Burgess & Douglas, 1999); or the infliction of pain and victim's suffering (e.g. Seto & Kuban, 1996). Thus, as noted by Marshall and Kennedy (2003), the dispute does not revolve around the range of typical behaviors enacted by sadists but rather around what constitutes the key element that elicits their sexual excitement.

Definitional and diagnostic challenges have led to differing levels of agreement among professionals when assessing sexual sadism, ultimately impacting on the ability to diagnose sexual sadism reliably (Nitschke, Mokros, Osterheider, & Marshall, 2013). The prevalence of sexually sadistic behavior (not disorder!) in the population at large has been estimated at about 2% to 3% (Baur et al., 2016). According to current psychiatric classification, sexual sadism *disorder* requires not only extended duration of the condition (i.e., more than six months) but also one of two additional aspects: Either sadistic acts against a non-consenting individual or distress/impairment for the person afflicted (American Psychiatric Association, 2013). In samples of sexual offenders the prevalence of sexual

sadism disorder has been estimated somewhat higher, with percentages ranging up to 10% of rapists (Eher et al., 2016). Finally, among the perpetrators of sexual homicide, the prevalence of sexual sadism disorder has been reported at about one-third in samples from Germany (Briken, Habermann, Berner & Hill, 2005: 36.7%,  $N = 166$ ) and the US (Geberth & Turco, 1997: 29.3%,  $N = 232$ ). Given the general uncertainty of clinical diagnoses in the forensic domain (Mokros, Habermeyer, & Küchenhoff, 2018) and in light of the variability of observer agreement on sexual sadism in particular (Nitschke et al., 2013), it remains unclear how high the prevalence of sexual sadism truly is among the perpetrators of sexual homicide. While sadism was recently shown to be less relevant for offense recidivism than customary indicators of risk (such as antisocial personality or psychopathy; Eher et al., 2016, Study 2), it might still be the case that sexual sadism is a primary force behind committing sexual offense that are rare, but most severe (i.e., sexual homicide). A meta-analysis of seven studies with a total sample comprising 2,169 individuals (Eher et al., 2016, Study 1) showed that sexual sadism was associated with a slightly higher risk of sexual recidivism (risk ratio = 1.38), yet not to a statistically significant degree ( $p = .052$ ).

The current study was planned to assess the psychometric properties and test the applicability of an IRT model (two-parameter logistic model, 2PLM, aka Birnbaum model) for a behavioral index of sexual sadism, the Sexual Sadism Scale (SeSaS; Mokros, Schilling, Weiss, Nitschke, & Eher, 2014). Secondly, the current study was meant to yield a robust estimate of the prevalence of sexual sadism among the perpetrators of sexual homicide offenses. For this purpose, a method was used that allows gauging prevalence from the minimum level of the latent trait associated with a given cut-off, the Posterior Probability of Diagnosis (PPOD) index (Lindhiem, Kolko, & Yu, 2013).

The SeSaS (Mokros et al., 2014) is a checklist of dichotomous (yes/no) items that consists of two parts: Part 1 contains 11 items that code for crime scene behavior, including

aspects like the gratuitous exertion of violence or confinement of the victim. These behavioral indicators were derived empirically from a larger pool of items showing content validity according to a survey of experts in the area of sexual sadism (Marshall, Kennedy, Yates, & Serran, 2002). Part 2 of the SeSaS instrument comprises three biographical items (planful conduct, prior sadistic acts beyond listed offenses, and arousability through sadistic fantasies or acts). The composite score of the Part 1 items was shown to have excellent interrater agreement (intra class correlation coefficient,  $ICC_{[2,5]}$ , i.e., average measure, absolute agreement = .91) in a sample of 20 cases assessed by five raters (Mokros et al., 2014). The Part 1 sum score showed a moderate to substantial correlation with clinical diagnoses of sadism ( $r_{pc} = .55$  according to Eher et al., 2016;  $r_{pc} = .46$  according to Longpré, Proulx, & Brouillette-Alarie, 2018;  $AUC = .87$  according to Mauzaite, Sauter, Seewald, & Dahle, 2017). Furthermore, the Part 1 sum score was strongly correlated ( $r = .66$ ) with the Massachusetts Treatment Center Sadism Scale (MTCSS; Longpré, Guay, & Knight, 2017).

For the predecessor of the SeSaS Part 1, a cut-off score of 4 points has been suggested as being indicative of sexual sadism (Nitschke, Osterheider, & Mokros, 2009). Across four samples of male (84.8%) and female offenders (15.2%) from Germany and the US (total  $N = 591$ ), the overall sensitivity of the cut-off with regards to a diagnosis of sexual sadism according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision (*DSM-IV-TR*; American Psychiatric Association, 2000) was estimated at 95%, the specificity at 99% (Nitschke et al., 2013). A prior study on the items now forming Part 1 of the SeSaS instrument showed good absolute model fit for a one-factorial structure in a confirmatory factor analysis (Root Mean Square Error of Approximation [RMSEA] = .05), even though the incremental fit index (here: Comparative Fit Index, CFI) was below commonly accepted standards (.89; Mokros, Schilling, Eher, & Nitschke, 2012). Finally, previous analyses yielded support for scalability of the Part 1 items in terms of non-metric

item response theory (Nitschke et al., 2009) or the 1-parameter logistic (aka Rasch) model (Mokros et al., 2012). Note that the SeSaS instrument was developed into a structured professional judgment instrument with more detailed item descriptions subsequently (Mokros et al., 2014).

### **Method**

The sample used in the study comprised of 350 male sexual killers, who perpetrated against female victims aged 14 years<sup>1</sup> or above and served a custodial sentence within HM Prison Service in England and Wales. Homicides were non-serial, with majority of offenders killing a single victim, and six cases having two victims (with the maximum timeframe between killing the two victims established as three hours). The criteria for sexual homicide included offenses where a sexual element in the killing was evidenced, suspected or admitted. The sample represented a full data search of all cases stored electronically in the Offender Assessment System in England and Wales captured from the beginning of its existence in the early 2000's (i.e. from that date, the offender was still serving a prison sentence). The actual time frame of the index offenses committed by the perpetrators ranged from the 1950's to 2010's. Details of the offence events were collected from the Public Protection Unit Database (PPUD).

The analyses reported herein were focused on the 11 dichotomous indicator variables of the SeSaS coding for crime-scene behavior (i.e., the SeSaS Part 1). The presumed unidimensional structure of these 11 items was assessed by confirmatory factor analysis (CFA) using the program Mplus, version 6.12 for Mac (Muthén & Muthén, Los Angeles, CA). A robust estimator that is suitable for categorical items was chosen for the CFA (i.e., weighted least squares means and variance adjusted; WLSMV). IRT analyses based on the

---

<sup>1</sup> The age of the victim was set at 14 years of age in order to offer consistency with previous research (Carter & Hollin, 2010).

two-parameter logistic model (2PLM), also known as Birnbaum model, were conducted in Mplus, version 6.12 for Mac, as well. The 2PLM was obtained through maximum likelihood estimation with robust standard errors (MLR).

The internal consistency of the SeSaS Part 1 items was assessed both at the factor level (in terms of MacDonald's omega,  $\omega$ ) and at the manifest level (in terms of the 2PLM reliability estimate for dichotomous data, *Rho*, developed by Dimitrov, 2003b). For  $\omega$ , a 95% bootstrap confidence interval (95% bCI) was obtained based on 1,000 bootstrap draws. *Rho* was estimated using the program IRT-TRUE by Dimitrov (2003a).

According to Bayes's theorem, the individual level on the latent trait being measured by the IRT model can be gauged through expected a posteriori (EAP) scores. For EAP estimates, a density distribution is obtained for the latent trait of an individual based on prior information (e.g., the individual response pattern); the expected value of said distribution is used as the person parameter of the person in question (Walter & Rost, 2011). Unlike maximum likelihood estimation, EAP estimates are also available for individuals for whom none of the items or all items were coded as present (Muraki & Engelhard, 1985).

Finally, the EAP person parameter estimates derived from the 2PLM were analyzed in terms of the Posterior Probability of Diagnosis (PPOD) index (Lindhiem et al., 2013). Due to the differential weighting of items in terms of their discrimination parameter within the 2PLM, the same total score may reflect different levels of the underlying trait, depending on the combination of items coded as present in a given case. Thus, all item profiles that occurred in the sample and equaled the cut-score of four points (e.g., 01110100000 or 11110000000) were identified; the minimum EAP person parameter associated with any of these profiles was determined; all individuals whose posterior probability of their EAP person parameters being equal to or above that minimum level (i.e., the PPOD index) was at least .5 (regardless of the actual sum score) were considered as tentatively diagnosed as

sadists based on the SeSaS; then, the agreement between those with manifest scores  $\geq 4$  and those with a PPOD index  $>.5$  was checked, also in terms of sensitivity and specificity. For the calculation of the PPOD index, Method A from Lindhiem et al. (2013) was used (i.e., based on a normal cumulative distribution function).

The research plan was reviewed by the National Research Committee and found to comply with ethical standards. Moreover, the authors complied with APA ethical standards in collecting, analyzing, and interpreting the data for the current study.

### Results

In order to establish inter-rater agreement of the instrument, 28 cases were independently blind-coded by two raters who are Chartered Forensic Psychologists with the British Psychological Society and are registered with the Health and Care Professions Council. The intra-class correlation coefficient ( $ICC_{[2, 1]}$ , i.e., single measure, absolute agreement) on the total score for Part 1 of the SeSaS calculated to .80 (95% CI [.58, .90]). Adopting the rules of thumb suggested by Cicchetti and Sparrow (1981; cf., Fleiss, 1981), the agreement on the individual items was excellent for three items (nos. 1, 5, and 8), good for another three items (nos. 3, 6, and 7), fair for two items (nos. 4 and 9), and poor for one item (no. 2; see Table 1 for  $\kappa$  estimates). Note that  $\kappa$  could not be computed for two individual items (no. 10 and 11) due to perfect agreement (joint absence).

The sample mean of the SeSaS Part 1 sum score was 2.67 ( $SD = 1.71$ ,  $Md = 2$ ), with values ranging from 0 to 10. The distribution of the SeSaS Part 1 total score was skewed to the right (skewness = 0.95, i.e., had a longer right tail), leptokurtic (kurtosis = 1.71), and unimodal (mode = 2). Out of 350 individuals, 94 (26.9%) had been assigned a SeSaS Part 1 sum score of 4 or above.

Within CFA, a single-factor solution with 22 free parameters had the following model fit properties: CFI = .87, RMSEA = .054 (90% CI [.037, .070]), and  $\chi^2(44) = 88.70$  ( $p <$

.001). Thus, the absolute fit index (RMSEA) was indicative of good fit ( $<.05$ ), whereas the incremental fit index (CFI) was below the commonly accepted standard for good fit of  $.95$ . In the present case, the magnitude of the CFI is likely not informative, however, because it critically hinges on the suitability of the null (or baseline) model.<sup>2</sup> Therefore, the 90% CI for the RMSEA coefficient (i.e.,  $[.037, .070]$ ) is more informative presently and indicates good model fit. The fully standardized factor loadings ranged from  $.29$  (item 7) to  $.80$  (item 3), all  $p < .01$  (two-sided). For the latent factor based on the Part 1 items,  $\omega$  was calculated at  $.84$  (95% bCI  $[.80, .89]$ ). Hence, the internal consistency of the Part 1 items was good.

The two-parameter logistic model (2PLM) comprises 22 free parameters. Estimating the model with the current data yielded a log-likelihood of  $-1381.87$ . The corresponding value for the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) were  $2807.73$  and  $2892.60$ , respectively. Model fit was tested through bivariate item comparisons. There were 3 occurrences of significant item misfit ( $|z| > 1.96$ ) among 220 bivariate item comparisons ( $11 \times 10/2 = 55$  non-redundant item pairs with 4 possible numerical codings each, namely 0/0, 0/1, 1/0, and 1/1). Given that one should expect 5% (i.e., about 11 such violations) under a Type I error rate of  $.05$ , the observed rate points toward superior goodness of fit ( $p = .999$ ) in a cumulative binomial test. Moreover, there was not a single occurrence of significant misfit for the 22 univariate item fit statistics ( $11 \text{ items} \times 2$  possible codings  $[0/1]$ ).

Table 1 shows the item parameters (discrimination,  $a_i$ , and difficulty,  $b_i$ ) for the 11 items of the SeSaS Part 1 according to the 2PLM, along with the corrected polychoric part-whole correlations. At 3.14, item 3 (torture) had the highest discrimination parameter ( $a_i$ )

---

<sup>2</sup> As Kenny (2015) points out the CFI should not be computed if the RMSEA of the null model is smaller than  $.158$  (cf., Rigdon, 1996). For the data at hand, the RMSEA of the null model is  $.133$ . This means that the null model (without any inter-correlations) already describes the data quite well. Consequently, there is little to be gained in terms of an incremental fit index such as the CFI or the Tucker Lewis Index (Kenny, 2015).

estimate. Thus, item 3 afforded the maximum of information on the latent trait ( $\theta$ ) of sexual sadism within the sample. The lowest  $a_i$  estimate was 0.43 (for item 7, excessive violence). Consequently, excessive violence does not distinguish well among those with lower or higher levels of  $\theta$ . Thus, except for item 7, none of the items had an  $a_i$  estimate below the minimum value of 0.5 usually observed in 2PLM applications (Reeve & Fayers, 2005). The different gradients of the  $a_i$  estimates are reflected by the slopes of the item characteristic curves in Figure 1 (Panel a) within the online supplemental material, with higher  $a_i$  values equaling steeper slopes. Item 3 (torture) yields the maximum information (see online supplemental material, Figure 1, Panel b) but only differentiates within a narrow spectrum of the latent trait  $\theta$ .

The estimates for item difficulty ( $b_i$ ) ranged from -2.46 (for item 1, sexual arousal) up to 3.96 (for item 11, taking trophies/keeping records). Thus, most  $b_i$  estimates were in the range from -3 to 3 commonly encountered in 2PLM modelling, with the exception of one item (no. 11). Looking at the test information function (see online supplemental material, Figure 1, Panel c) it becomes clear that the maximum total information is conveyed at  $\theta = 1.39$ . Thus, within the sample analyzed, the SeSaS items conveyed most information at an elevated trait level which is similar (in terms of difficulty) to items 3 (torture), 4 (degradation/humiliation), and 9 (ritualistic behavior), with  $b_i$  estimates of 1.34, 1.39, and 1.55, respectively.

Looking at the test characteristic curve (see online supplemental material, Figure 1, Panel d) it becomes clear, that the association between the latent trait,  $\theta$ , and the expected score is most reliable at medium to high trait levels (i.e., for  $\theta > 0$ ). This is concomitant with the focus on the severe (or forensically relevant) variant of sexual sadism. A global 2PLM estimate of scale reliability ( $Rho$ ) was estimated at .76, somewhat lower than  $\omega$  (.84).

A kernel density estimate for the distribution of the EAP person parameter estimates is provided as Figure 2 within the online supplemental material. The distribution is bimodal, with a local maximum at approximately -1 and a global maximum at 0. For each possible manifest score on the SeSaS Part 1 ( $s_k$ ), we checked the minimum and maximum  $\theta$  levels associated with profiles affording the total score in question within the sample (i.e.,  $\min[\hat{\theta}|s_k]$  and  $\max[\hat{\theta}|s_k]$  for  $k = 1..12$ ; see Table 2 within the online supplemental material). For a SeSaS Part 1 total score of 1, for instance, the minimum  $\theta$  level associated with this manifest score was estimated at -1.10, whereas  $\max(\hat{\theta}|1)$  was estimated at -0.37. For the cut-score of 4 points recommended for the SeSaS Part 1, the minimum estimate was 0.14.

Next, we calculated the PPOD index, that is, the posterior probability of having a  $\theta$  level of .14 or higher with any pattern of items coded as present within the sample. When the items 1, 2, and 8 were coded as present,  $\theta$  was estimated at .178, for example. Thus, a case with the item profile of 11000001000 surpassed the latent trait level minimally implied by the manifest cut-score of 4. Finally, all individuals whose PPOD index (i.e., the posterior probability of their  $\hat{\theta}_j$  being at least .14, given their item profile) was .5 or higher were assigned to the  $\text{PPOD} \geq .5$  group.

The cell entries for a  $2 \times 2$  contingency table (SeSaS Part 1 total score  $\geq 4$  / PPOD  $\geq .5$ : no / no, no / yes, yes / no, and yes / yes) were 218, 38, 1, and 93, respectively. In other words, only a single individual would not be considered sadistic based on the PPOD index, but regarded as sadistic based on the SeSaS Part 1 sum score. Vice versa, 38 individuals had PPOD index values indicative of sadism despite SeSaS Part 1 sum scores  $<4$ . Nevertheless, the agreement between the two modes of assessment was high ( $z = 10.78$ ,  $p < .001$ ).

Based on the PPOD index  $>.5$  criterion, sensitivity of the SeSaS Part 1 cut-score (4 points) was estimated at 71.0% (95% CI [62.4%, 78.6%]), whereas the specificity was

estimated at 99.5% (95% CI [97.5%, 100%]). Consequently, the cut-score of 4 points is relatively conservative, maximizing specificity rather than sensitivity.

Furthermore, based on the PPOD index  $>.5$  criterion, the prevalence of sexual sadism was estimated at 37.4% (95% CI [32.0%, 42.7%]). Hence, at least one-third of sexual homicide offenders can be expected to be sexual sadists. Finally, the correlation between the SeSaS Part 1 total score and a dichotomous variable coding for whether the offender at hand had been diagnosed as a sadist according to his files at some point was  $r_{pb} = .57$  ( $p < .001$ ). Despite the variability of the methods and criteria used by clinicians and expert witnesses to reach such a diagnosis, the strength of the association attests to the criterion validity of the SeSaS for sexual homicide offenders.

### Discussion

The current study assessed sexual sadism in a nationwide sample of men who had committed sexual homicide offenses in England and Wales. Using CFA, the factorial structure of a file-based assessment instrument for sexual sadism, the SeSaS, was corroborated. Focusing on the items that code for crime-scene behavior (i.e., Part 1 of the SeSaS), both inter-rater agreement and internal consistency could be ascertained. Moreover, the corresponding sum score was shown to be associated with clinical diagnoses of sadism derived from the files.

Within the framework of IRT, the SeSaS Part 1 items were concomitant with the 2PLM. That is, the association between the latent trait of sexual sadism and the occurrence of behavioral indicators could be described by logistic functions with two parameters (discrimination and difficulty). The results of the 2PLM modeling imply that the SeSaS captures moderate to severe levels of the latent trait of sexual sadism. This extends earlier research on selective (Nitschke et al., 2009) or smaller samples (Mokros et al., 2012) testing non-metric IRT or Rasch models for the predecessor of the SeSaS Part 1, respectively.

More specifically, the 2PLM model opens up the possibility of assessing the uncertainty at the latent trait level that is associated with recommended cut-offs at the observed level. In the case of the SeSaS, the recommended cut-score of 4 points represents a conservative threshold, compared with the so-called Posterior Probability of Diagnosis (PPOD) index. Moreover, the prevalence of sexual sadism was estimated based on the PPOD index. According to the 95% CI of the prevalence estimate, at least one-third of sexual homicide offenders are sexual sadists. This estimate accords well with earlier findings at the manifest level (Briken et al., 2005; Geberth & Turco, 1997). In applying the SeSaS Part 1 items it should be noted, however, that the diagnostic usefulness of item 11 and 8 may be limited due to very high estimates of the difficulty parameters. Put differently, these items only concern a very minor fraction of individuals. Similar reservations apply to item 2, yet for another reason (i.e., sub-optimal inter-rater agreement).

Recently, Eher et al. (2016) showed that the *DSM-IV-TR* diagnosis of sexual sadism was only moderately related to violent or sexual offense recidivism. In addition, Eher and colleagues demonstrated that the diagnosis of sexual sadism does not add incremental validity for assessing the risk of re-offending once customary risk factors like antisocial personality or psychopathy have been controlled for. The current results, however, show that sexual sadism is a relevant condition in the most grievous (i.e., lethal) forms of sexual aggression. Therefore, it might turn out in further studies, that the SeSaS - although not predictive for general violent reoffense (Eher et al., 2016) - might contribute to predicting at least most grievous forms of (sexual) aggression. Berner, Hill, and Briken (2018) emphasized the importance of diagnosing sexual sadism reliably with regard to treatment planning. We might add that delineating the diagnosis based on behavioral indicators (cf., Kingston & Yates, 2008) is particularly important with individuals who are likely motivated to deny or downplay sadistic urges or fantasies, such as sexual homicide offenders. Furthermore, an

operational definition of sexually sadistic conduct against non-consenting individuals (as provided by the SeSaS) could further our understanding of the commonalities and differences with sadomasochistic roleplay.

The current study dealt with male non-serial offenders who perpetrated against females 14 years old and over. This selection criteria allowed 1) to focus on the most prevalent group of offenders in the correctional facilities that will most likely require the SeSaS assessment (Proulx et al., 2007); 2) examine a complex group given the lack of victim statements available and limited evidence behavioral patterns due to the non-serial nature of the offence and 3) investigate a group more likely eligible for parole than serial homicide offenders. However, such sample restrictions also create limitations as the results are applicable to the type of offenders included only. Thus, the present research does not generalize to the rare group of serial sexual homicide for whom the prevalence of sexual sadism is presumably even higher (Warren et al., 1996). Moreover, testing the criterion validity of instruments like the SeSaS with clinical diagnoses of sadism is a somewhat suboptimal strategy given the concerns about the reliability of clinician judgments for this diagnosis (Nitschke et al., 2013). Therefore, physiological measurement may provide further evidence of criterion validity (see, e.g., Seto, Lalumière, Harris, & Chivers, 2012, for a useful stimulus set) even though extant results using phallometry yielded nil correlations (Longpré et al., 2018).

In sum, the analyses presented herein confirm the appropriateness of the SeSaS as an assessment instrument for forensically relevant sexual sadism in English and Welsh offenders, extending its validity beyond the US, Canadian, and German samples scrutinized so far.

## References

- American Psychiatric, A. (2000). *Diagnostic and statistical manual of mental disorders (4<sup>th</sup> ed.)*. Washington, DC: Author.
- Baur, E., Forsman, M., Santtila, P., Johansson, A., Sandnabba, K., & Långström, N. (2016). Paraphilic sexual interests and sexually coercive behavior: A population-based twin study. *Archives of Sexual Behavior, 45*, 1163–1172. doi:10.1007/s10508-015-0674-2
- Berner, W., Hill, A., & Briken, P. (2018). Sexual sadism: Pharmacological and psychotherapeutic approaches. In J. Proulx, E. Beauregard, A. J. Carter, A. Mokros, R. Darjee, & J. James (Eds.), *Routledge international handbook of sexual homicide studies* (pp. 317-325). London, UK: Routledge
- Briken, P., Habermann, N., Berner, W., & Hill, A. (2005). The influence of brain abnormalities on psychosocial development, criminal history and paraphilias in sexual murderers. *Journal of Forensic Sciences, 50*, 1204-1208. doi:10.1520/JFS2004472
- Carter, A. J., & Hollin, C. R. (2010). Characteristics of non-serial sexual homicide offenders: A review. *Psychology, Crime & Law, 16*, 25-45. doi: 10.1080/10683160802621933
- Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127–137.
- Eher, R., Schilling, F., Hansmann, B., Pumberger, T., Nitschke, J., Habermeyer, E., & Mokros, A. (2016). Sexual sadism and violent reoffending. *Sexual Abuse: A Journal of Research and Treatment, 28*, 46-72. doi:10.1177/1079063214566715
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions (2<sup>nd</sup> ed)*. New York: Wiley.
- Dimitrov, D. M. (2003a). *IRT-TRUE: A computer program for estimation of true-score parameters and reliability from IRT estimates of item parameters*. Fairfax, VA: George Mason University, Graduate School of Education.

- Dimitrov, D. M. (2003b). Marginal true-score measures and reliability for binary items as a function of their IRT parameters. *Applied Psychological Measurement, 27*, 440-458.
- Geberth, V. J. & Turco, R. N. (1997). Antisocial personality disorder, sexual sadism, malignant narcissism, and serial murder. *Journal of Forensic Sciences, 42*, 49-60.  
doi:10.1520/JFS14067J
- Kenny, D. A. (2015). *Measuring model fit*. Retrieved from: <http://davidakenny.net/cm/fit.htm>
- Kingston, D. A., & Yates, P. M. (2008). Sexual sadism: Assessment and treatment. In D. R. Laws & W. O'Donohue (Eds.), *Sexual deviance: Theory, assessment, and treatment* (2<sup>nd</sup> ed., pp. 231–249). New York, NY: Guilford Press.
- Lindhiem, O., Kolko, D. J., & Yu, L. (2013). Quantifying diagnostic uncertainty using item response theory: The Posterior Probability of Diagnosis Index. *Psychological Assessment, 25*, 456-466. doi:10.1037/a0031392
- Longpré, N., Guay, J.-P., & Knight, R. A. (2017). MTC Sadism Scale: Toward a dimensional assessment of severe sexual sadism with behavioral markers. *Assessment*. Advance online publication. doi:10.1177/1073191117737377
- Longpré, N., Proulx, J., & Brouillette-Alarie, S. (2018). Convergent validity of three measures of sexual sadism: Value of a dimensional measure. *Sexual Abuse: A Journal of Research and Treatment, 30*, 192-208.. doi:10.1177/1079063216649592
- MacCulloch, M. J., Snowden, P. R., Wood, P. J., & Mills, H. E. (1983). Sadistic fantasy, sadistic behaviour and offending. *British Journal of Psychiatry, 143*, 20-29.  
doi:10.1192/bjp.143.1.20
- Marshall, W. L., & Kennedy, P. (2003). Sexual sadism in sexual offenders. An elusive diagnosis. *Aggression and Violent Behavior, 8*, 1-22. doi:10.1016/S1359-1789(01)00052-0

- Marshall, W. L., Kennedy, P., Yates, P., & Serran, G. (2002). Diagnosing sexual sadism in sexual offenders: Reliability across diagnosticians. *International Journal of Offender Therapy and Comparative Criminology*, *46*, 668-677.  
doi:10.1177/0306624X02238161
- Mauzaite, A., Sauter, J., Seewald, K., & Dahle, K. P. (2017, September). *Tatbildbasierte Screening-Instrumente für die Diagnosen Pädophilie und Sexuellen Sadismus in der Gruppe der Hochrisikotäter* [Crime-scene based screening instruments for the diagnoses of pedophilia and sexual sadism within the group of high-risk offenders]. Paper presented at the 17<sup>th</sup> Meeting of the Chapter on Legal Psychology of the German Psychological Society (DGPs), Jena, Germany.
- Mokros, A., Habermeyer, E., & Küchenhoff, H. (2018). The uncertainty of psychological and psychiatric diagnosis. *Psychological Assessment*, *30*, 556-560.  
doi:10.1037/pas0000524
- Mokros, A., Schilling, F., Eher, R., & Nitschke, J. (2012). The Severe Sexual Sadism Scale: Cross-validation and scale properties. *Psychological Assessment*, *24*, 764-769.  
doi:10.1037/a0026419
- Mokros, A., Schilling, F., Weiss, K., Nitschke, J., & Eher, R. (2014). Sadism in sexual offenders: Evidence for dimensionality. *Psychological Assessment*, *26*, 138-147. doi:10.1037/a0034861
- Muraki, E., & Engelhard, G. (1985). Full-information item factor analysis: Applications of EAP scores. *Applied Psychological Measurement*, *9*, 417-430.  
doi:10.1177/014662168500900411
- Myers, W. C., Burgess, A. W., Burgess, A. G., & Douglas, J. E. (1999). Serial murder and sexual homicide. In V. van Hasselt & M. Herson (Eds.), *Handbook of psychological approaches with violent offenders* (pp. 153-172). New York: Springer.

- Nitschke, J., Mokros, A., Osterheider, M., & Marshall, W. L. (2013). Sexual sadism: Current diagnostic vagueness and the benefit of behavioral definitions. *International Journal of Offender Therapy and Comparative Criminology*, *57*, 1441-1453.
- Nitschke, J., Osterheider, M., & Mokros, A. (2009). A cumulative scale of severe sexual sadism. *Sexual Abuse: A journal of Research and Treatment*, *21*, 262-278.
- Proulx, J., Cusson, M., & Beauregard, E. (2007). Sexual murder: Definitions, epidemiology and theories. In J. Proulx, E. Beauregard, M. Cusson, & A. Nicole (Eds.), *Sexual murderers: A comparative analysis and new perspectives*. (pp. 9-28). Chichester, UK: John Wiley & Sons Ltd.
- Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. In P. Fayers & R. Hays (Eds.), *Assessing quality of life in clinical trials: Methods and practice* (2<sup>nd</sup> ed., pp. 55-73). Oxford, UK: Oxford University Press.
- Ressler, R. K., Burgess, A. W., & Douglas, J. E. (1988). *Sexual homicide: Patterns and motives*. New York: Lexington.
- Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling*, *3*, 369-379.  
doi:10.1080/10705519609540052
- Seto, M. C., & Kuban, M. (1996). Criterion-related validity of a phallometric test for paraphilic rape and sadism. *Behaviour Research & Therapy*, *34*, 175-183.  
doi:10.1016/0005-7967(95)00056-9
- Seto, M. C., Lalumière, M. L., Harris, G. T., & Chivers, M. L. (2012). The sexual responses of sexual sadists. *Journal of Abnormal Psychology*, *121*, 739-753.  
doi:10.1037/a0028714

von Krafft-Ebing, R. (1906). *Psychopathia sexualis, with especial reference to the antipathic sexual instinct: A medico-forensic study* (12<sup>th</sup> ed.; F. J. Rebman, Trans.). New York,

NY: Rebman. Retrieved from <https://archive.org/details/psychopathiasexu00kraffuoft>

Walter, O., & Rost, J. (2011). Psychometrische Grundlagen von Large Scale Assessments [Psychometric foundations of large-scale assessments]. In L. F. Hornke, M. Amelang, & M. Kersting (Eds.), *Methoden der psychologischen Diagnostik* (pp. 88-150).

Göttingen, Germany: Hogrefe.

Warren, J. I., Hazelwood, R. R., & Dietz, P. E. (1996). The sexually sadistic serial killer.

*Journal of Forensic Sciences, 41*, 970-974.

Table 1

*Corrected Polychoric Item Part-Whole Correlations ( $r_{pc}$ ), Estimates of Item Parameters According to the Two-Parameter Logistic Model ( $N = 350$ ), Cohen's  $\kappa$  Coefficients, and Percentage of Items Coded as Present (%).*

No	SeSaS Part 1 Items	$r_{pc}$ ( $p$ , two-sided)	Discrimina tion ( $a_i$ )	Difficulty ( $b_i$ )	Cohen's $\kappa$	%
1.	Sexual arousal during the crime scene behaviors	.22 (.009)	.889	-2.458	.78	87.1
2.	Exertion of power, control, or dominance	.54 (<.001)	2.134	-0.408	.39	62.6
3.	Torturing the victim	.65 (<.001)	3.142	1.344	.65	12.0
4.	Degrading or humiliating behavior directed toward the victim	.59 (<.001)	2.249	1.385	.51	13.7
5.	Mutilation of sexual areas of the victim's body	.58 (<.001)	1.175	1.911	1.0	14.0
6.	Mutilation of other parts of the victim's body	.51 (<.001)	1.116	2.652	.65	7.7
7.	Excessive physical violence	.19 (.004)	0.434	1.813	.71	32.0
8.	Insertion of objects into the victim's bodily orifices	.33 (<.001)	0.701	2.939	.87	13.1
9.	Ritualistic behavior	.48 (<.001)	1.348	1.553	.42	16.9
10.	Confinement of the victim	.57 (<.001)	1.891	2.702	- *	2.6
11.	Taking trophies	.29 (<.01)	0.827	3.955	- *	4.9

*Note.* The Sexual Sadism Scale (SeSaS) is from Mokros et al. (2014).

\*  $\kappa$  could not be computed due to perfect agreement (joint absence).