

An Evaluation of the Reliability, Construct Validity, and Factor Structure of the Static-2002R

International Journal of
Offender Therapy and
Comparative Criminology
1–24

© The Author(s) 2015

Reprints and permissions:

sagepub.com/journalsPermissions.nav

DOI: 10.1177/0306624X15595228

ijo.sagepub.com



**Sandy Jung¹, Liam Ennis², Chantal A. Hermann³,
Anna T. Pham³, Alberto L. Choy^{4,5}, Gabriela Corabian⁶,
and Tarah Hook⁵**

Abstract

The fundamental psychometric properties of the subscales found in the Static-2002R, an actuarial measure of sexual recidivism risk, were evaluated in the current study. Namely, the reliability, concurrent and construct validity, and factor structure of the Static-2002R subscales were examined with a sample of 372 adult male sex offenders. In addition to using validated measures of sexual violence risk to examine concurrent validity, construct-related measures taken from extant risk measures and psychometric tests were correlated with three of the subscales to assess overall construct validity. Moderate support was found for the reliability of the Static-2002R. The concurrent and construct validity of the General Criminality, Persistence of Sexual Offending, and Deviant Sexual Interest subscales were supported. Generally, these findings further support the Static-2002R as a valid sex offender risk appraisal instrument that encompasses multiple distinct, clinically relevant, risk domains.

Keywords

Static-2002R, sex offenders, construct validity, risk assessment, reliability

¹MacEwan University, Edmonton, Alberta, Canada

²Integrated Threat and Risk Assessment Centre, Edmonton, Alberta, Canada

³Carleton University, Ottawa, Ontario, Canada

⁴University of Alberta, Edmonton, Alberta, Canada

⁵Alberta Hospital Edmonton, Edmonton, Alberta, Canada

⁶University of Saskatchewan, Saskatoon, Saskatchewan, Canada

Corresponding Author:

Sandy Jung, Department of Psychology, MacEwan University, P.O. Box 1796, Edmonton, Alberta, Canada T5J 2P2.

Email: sandy.jung@macewan.ca

The past two decades have witnessed the introduction, validation, and widespread acceptance of actuarial procedures for the purpose of estimating risk for future violence. The popularity of actuarial risk measures can be attributed to their ease of administration and relative superiority over structured professional judgment (SPJ) approaches in terms of inter-rater reliability and predictive accuracy (Archer, Buffington-Vollum, Stredny, & Handel, 2006; Hanson & Morton-Bourgon, 2009). Since its introduction in 1999, the Static-99 (Hanson & Thornton, 1999) has become the most extensively researched and widely used risk assessment instrument for sexual offenders (Archer et al., 2006; McGrath, Cumming, Burchard, Zeoli, & Ellerby, 2010). Although the Static-99R is a reliable and valid measure of risk of sexual recidivism (Hanson, Lunetta, Phenix, Neeley, & Epperson, 2014), it is limited in that it is comprised exclusively of static, historical variables. This is problematic as it means the Static-99R provides risk assessors with little information regarding the individual and contextual factors that underlie a particular offender's propensity for sex offending. Thus, the Static-99R does not contribute to an explanation of *why* an individual may be inclined to engage in sexually abusive behavior and does not identify factors that need to be addressed to mitigate risk.

Recent years have witnessed a call for the development of risk assessment instruments that, unlike the first generation of actuarial assessment instruments, are composed of items that reflect clinically meaningful constructs (Hanson, 2009; Mann, Hanson, & Thornton, 2010). With the goals of improving predictive accuracy, clarifying scoring, and most importantly, integrating risk-relevant constructs into the actuarial scheme, the Static-2002 was introduced (Hanson & Thornton, 2003) and subsequently revised (Static-2002R; Helmus, Thornton, Hanson, & Babchishin, 2012). Although not as commonly used as the Static-99, one of five sex offender programs in the United States employs the Static-2002 while as many as 40% of Canadian community programs use the measure (McGrath et al., 2010).

Like the Static-99, the Static-2002 was designed as a brief actuarial measure for the prediction of sexual recidivism based on information that is routinely available in police and criminal justice records. The aim was to retain the advantages of the Static-99, such that it should be applicable to a wide range of offenders and lend itself to easy and reliable scoring. Furthermore, the Static-2002 was designed to be more coherent and conceptually clear than the Static-99 with regard to the risk-relevant constructs that are being assessed, more consistent across scoring criteria, and demonstrate improved predictive accuracy (or at least retain the same levels as the Static-99; Hanson & Thornton, 2003). In support of this revision, research has found that the Static-2002/R is equally predictive of sexual, violent, and general recidivism as the Static-99R (Bengtson & Langstrom, 2007; Hanson, Helmus, & Thornton, 2010; Hanson & Thornton, 2003).

Establishing predictive validity is critical for all risk assessment measures. In addition to this, there is also a need to investigate a measure's psychometric properties and to assess whether these properties are consistent across studies using samples independent from those used to develop the measure (American Educational Research Association [AERA], American Psychological Association [APA], & National Council

on Measurement in Education [NCME], 1999; Jung, Daniels, Friesen, & Ledi, 2012; Nunes & Babchishin, 2011). With regard to reliability, internal consistency estimates of the Static-2002R are generally adequate and appropriate (e.g., Cronbach's α of .68; Langton, Barbaree, Hansen, Harkins, & Peacock, 2007; one would not expect high internal consistency as items are selected to independently predict recidivism and do not measure the same underlying construct), and indices of inter-rater reliability are generally high for individual items (Langton, Barbaree, Hansen, et al., 2007).

The Static-2002 and its revision, Static-2002R, have also demonstrated good concurrent validity. Langton, Barbaree, Hansen, et al. (2007) investigated the concurrent validity of the Static-2002 and found that it was moderate to good. Correlations between total scores for the Static-2002 and other measures of risk of sexual recidivism ranged between .54 and .80, the highest correlation observed was with the Static-99. The same study investigated the construct validity of the Static-2002 by correlating the domains of the Static-2002 with theoretically relevant constructs derived from other measures associated with risk (i.e., total and factors scores for the Psychopathy Checklist-Revised [PCL-R] and selected items from the Sexual Violence Risk-20 [SVR-20] and Sex Offender Risk Appraisal Guide [SORAG]). The General Criminality content area was significantly correlated with all other measures of risk of sexual (e.g., Static-99, SORAG) and violent (e.g., Violence Risk Appraisal Guide [VRAG]) recidivism. Both the Persistence of Sexual Offending and Deviant Sexual Interests content areas significantly correlated with the variable concerning attitudes toward sexual offending. Langton and his colleagues also investigated the factor structure of the Static-2002. Their results indicated a five-factor model fit their data best. Consistent with the conceptual structure of the Static-2002, the first factor consisted of four items comprising the General Criminality content area. The second factor consisted of the three Persistence of Sexual Offending items. The third factor consisted of two of the three Deviant Sexual Interests content area items as well as the Unrelated Victims item and a negative loading for the Age at Release item. The fourth factor consisted of the two Relationship to Victims content area items and the Age at Release item. The last factor was composed of the convictions for Noncontact Sex Offenses item along with the negatively loaded Age at Release and Juvenile Arrest for Sexual Offense items.

A second study by Langton, Barbaree, Seto, et al. (2007) compared the Static-2002 with five other risk measures of sexual and/or violent recidivism (VRAG, SORAG, Rapid Risk Assessment for Sex Offense Recidivism [RRASOR], Static-99, Minnesota Sex Offender Screen Tool-Revised [MnSOST-R]) and found that the Static-2002 was significantly and positively correlated with each of the measures. All measures demonstrated predictive validity but the Static-99 and Static-2002 were superior in this regard. These findings demonstrate that the content areas of the Static-2002 have construct and concurrent validity using clinician-rated items from existing risk assessment measures and that the Static-2002 is a superior measure of risk for sexual recidivism. Langton and his colleagues examined the construct validity of the Static-2002 using risk tools, but given the overlap among most risk assessment measures, it would strengthen the construct validity of the Static-2002, particularly the revised version, Static-2002R, if construct-relevant variables from non-risk measures were used.

The purpose of the present study was to further explore the psychometric properties of the Static-2002R using measures that include, but were not limited to, extant risk measures. Specifically, the inter-rater reliability, concurrent and construct validity, and the factor structure of the Static-2002R were examined in this study. The reliability and validity of the Static-2002R total score and of the following content areas were investigated: Persistence of Sex Offending (PSO), Deviant Sexual Interests (DSI), and General Criminality (GC) subscales. The Relationship to Victims content area was excluded from the construct validity analyses as it is unknown what underlying construct the victim items are intended to measure. Similarly, the construct validity of age was also not examined given that it is a single item. Concurrent validity was investigated by correlating the Static-2002R with other measures of risk of sexual recidivism (Static-99R, a modified SORAG, and an adapted score based on the SVR-20 items). Construct validity, specifically convergent validity (i.e., measure/subscale is related to theoretically similar constructs), was investigated by correlating Static-2002R with measures of theoretically similar constructs obtained from file information, Personality Assessment Inventory (PAI), Multiphasic Sex Inventory (MSI-2), Sex History Questionnaire, Screening Scale of Pedophilic Interests, Childhood and Adolescent Taxon Scale, and Cormier–Lang criminal history indices, and selected items from the SVR-20.

Method

Participants

Archival data were gathered from patient files at an outpatient forensic psychiatric clinic and an affiliated secure forensic psychiatric hospital, both of which offer court-ordered assessment and treatment services for adjudicated offenders. Adult male sex offenders evaluated between 2001 and 2009 were identified and their patient files were coded for relevant variables. Of the total of 474 available files, 51 files did not include sufficient information to score the Static-2002R (e.g., no official criminal record), 50 offenders solely committed non-contact child pornography offenses, and one was a juvenile offender. Of the 372 offenders, 342 had calculable total scores for the Static-2002R. The mean age at the time of release for the remaining sample was 39.41 years ($SD = 13.58$) and ranged from 19 to 88 years. The average level of education of the sample was 11.36 years ($SD = 2.41$) with a range of 4 to 20 years. Regarding their index offenses only, 38.7% of subjects offended against child victim(s), 26.6% offended against adolescent victims—that is, between ages 13 to 17, 23.9% against adult victims, 4.8% had mixed-age victims, and victim age was unspecified for 5.9% of the sample. A majority of the sample offended against a female victim(s) in their index offense (94.1%), with 5.4% who solely offended against males, and 0.6% had both male and female victims in their index offense. About a fifth of the sample (18%) had prior convictions for sexual offenses on their record.

Measures

The current study used items and/or risk instruments and procedures, variables that were coded from files, and psychometric self-report inventories. Scores and coding were obtained from pre-existing archival files generated from inpatient and outpatient forensic facilities.

Risk items and totals. To examine the concurrent and construct validity of the Static-2002R, the Static-99R, SORAG, and SVR-20 were included, along with published measures of antisocial behavior and pedophilia used in the SORAG total score calculation.

Static-2002R. The Static-2002R contains 14 items grouped into five content areas (age at release, persistence of sexual offending, deviant sexual interests, relationship to victims, general criminality), and total scores can range from -2 to 13 (Helmus et al., 2012; Phenix, Doren, Helmus, Hanson, & Thornton, 2008). Inter-rater reliability has been shown to be high with an intra-class correlation (ICC) of .98 (Helmus & Hanson, 2007); however, the authors noted that this was exceptionally high and should not be considered representative of the typical circumstances in which the Static-2002 would be used. Modest internal consistency estimates were found for the content area subscales (Cronbach's α s for subscales ranged from .45 to .74 and for total score, .68; Langton, Barbaree, Hansen, et al., 2007). The Static-2002 has been shown to predict sexual, violent, and general recidivism with area under curve (AUCs) ranging from .64 to .79, and has been cross validated in several studies, often showing that it can outperform the Static-99 (Bengtson, 2008; Langton, Barbaree, Hansen, et al., 2007; Langton, Barbaree, Seto, et al., 2007; Looman & Abracen, 2010; Stalans, Hacker, & Talbot, 2010).

Static-99R. The Static-99R (Hanson & Thornton, 1999; Harris, Phenix, Hanson, & Thornton, 2003; Helmus et al., 2012) is a 10-item static risk assessment tool used to assess risk of sexual recidivism among adult males who have been charged with a sexual offense. The instrument includes 10 items, and total scores range from -3 to 12 . The Static-99 has demonstrated excellent inter-rater reliability (ICC = .98 in Rettenberger, Matthes, Boer, & Eher, 2010; ICC = .90 in Barbaree, Seto, Langton, & Peacock, 2001; ICC = .91, Langton, Barbaree, Hansen, et al., 2007), although one study found markedly lower values in an adversarial field setting (ICC = .64 in Murrie et al., 2009). The Static-99R also has good predictive validity for sexual, general violent, and general criminal recidivism (AUCs = .68, .70, and .72, respectively; Babchishin, Hanson, & Helmus, 2012).

SORAG. The SORAG (Quinsey, Harris, Rice, & Cormier, 2006) was developed to predict violent and sexual recidivism among adult male sexual offenders. The SORAG contains 14 static items that are weighted. The total score can range from -27

to 51, and individuals are assigned to one of nine risk categories (from 1 = *lowest* to 9 = *highest risk*). Researchers have reported excellent inter-rater reliability scores for the SORAG, with ICC values ranging from .93 to .96 (as reported in Harris, Rice, et al., 2003; Rettenberger et al., 2010). Research has demonstrated the predictive validity of the SORAG, with high accuracy for violent recidivism (AUCs ranged from .69 to .73) and moderate accuracy for sexual recidivism (AUCs ranged from .58 to .66; Bartosh, Garby, Lewis, & Gray, 2003; Harris, Rice, et al., 2003; Nunes, Firestone, Bradford, Greenberg, & Broom, 2002; Quinsey, Rice, & Harris, 1995; Rettenberger et al., 2010). Because phallometric test results and PCL-R scores were available in only a minority of cases, values for the Screening Scale for Pedophilic Interests (SSPI) and Childhood and Adolescent Taxon Scale (CATS) were substituted for calculating SORAG scores. Studies have shown that SORAG scores based on these substitutions retain their predictive validity (Quinsey et al., 2006).

SVR-20. The SVR-20 (Boer, Hart, Kropp, & Webster, 1997) is a structured professional guideline procedure developed to aid clinical judgment and prediction of sexual recidivism risk in adult sex offenders. Its 20 items are scored on a 3-point scale, grouped into three domains: psychosocial adjustment, sexual offenses, and future plans. Although the items can be summed to produce a total score as used in other empirical endeavors (e.g., de Vogel, de Ruiter, van Beek, & Mead, 2004; Langton, 2003; Parent, Guay, & Knight, 2011), it is important to note that this procedure was not endorsed by the authors of the SVR-20. Using this quantified scoring scheme, the total score for the SVR-20 could range from 0 to 40. Inter-rater reliability was good with an ICC of .75 (Parent et al., 2011). The predictive validity of the SVR-20 for sexual, general violent, and general criminal recidivism, using the total score, was considered good (AUCs = .77, .61, and .67, respectively; Rettenberger et al., 2010).

CATS. The CATS (Harris, Rice, & Quinsey, 1994) is an 8-item measure scored from file information, which has been used as an alternative index of psychopathy. The CATS was developed as a proxy measure for inclusion in the VRAG and SORAG in the absence of a PCL-R score (Quinsey et al., 2006). The CATS comprises items reflecting early life aggression and maladjustment that are associated with a trajectory of life-course persistent antisociality. The items are scored dichotomously and summed for a total score ranging from 0 to 8, and include elementary school maladjustment, teenage alcohol problems, childhood aggression rating, conduct disorder symptoms, suspension or expulsion, arrested under age of 16, parent alcoholism, lived with both biological parents to age 16. Predictive accuracy studies have shown almost identical results using the CATS as a replacement for the PCL-R (AUC = .75; Bartosh et al., 2003). The CATS has also demonstrated discriminant validity in a study with 106 male offenders, discriminating violent recidivists from other recidivists (*common language effect size* of 0.61; see Glover, Nicholson, Hemmati, Bernfeld, & Quinsey, 2002). The CATS has demonstrated good inter-rater reliability with a coefficient of .92, and good construct validity when correlated with PCL-R scores ($r = .89$; Glover et al., 2002).

SSPI. The SSPI (Seto & Lalumière, 2001) was developed to provide a measure of pedophilic sexual interests for clinical screening and research purposes, in cases where phallometric testing is unavailable (see Exhibit 8.2 of Quinsey et al., 2006; Seto, Harris, Rice, & Barbaree, 2004). The SSPI is a brief screening tool that summarizes the characteristics of a sex offender's victim(s) to identify offenders who are more likely to be pedophilic in their sexual interests. The scale is comprised of four items representing victim characteristics that are empirically established correlates of pedophilia (i.e., presence of male victim, more than one victim, victim is under the age of 12, unrelated victim). Total scores on the SSPI range from 0 to 5; having a male victim is assigned double the weight in scoring, this is because Seto and Lalumière (2001) found having a male victim explained nearly twice the variance as the other three items. Among adult male offenders with child victims, the SSPI was found to be significantly and positively associated with phallometric measured sexual arousal to children (Seto et al., 2004; Seto & Lalumière, 2001; Seto, Murphy, Page, & Ennis, 2003), and was significantly and positively correlated with violent recidivism (AUCs = .62-.67) and sexual recidivism (AUCs = .62-.69; the latter finding was significant for one of the two samples in the study; Seto et al., 2004).

Psychometric measures. Several self-report personality and sexual interest measures and their scales were used to examine the construct validity of the Static-2002R.

PAI. The PAI (Morey, 1991, 2007) is a commonly used self-report measure examining adult personality and psychopathology. It is comprised of 344 items that measure 22 non-overlapping scales, including 4 validity scales, 11 clinical scales, 2 interpersonal scales, and 5 treatment scales (Morey, 1991). PAI scales have been shown to have adequate to good internal consistency, ranging from $\alpha = .73$ to $.81$ (Boone, 1998; Boyle & Lennon, 1994; Morey, 1991) and adequate test-retest reliability, with median correlations ranging from $.73$ to $.83$ (Boyle & Lennon, 1994; Morey, 1991). PAI subscales related to malingering, violence, psychosis, personality disorder, and psychopathy have been validated in forensic settings (Douglas, Hart, & Kropp, 2001; Morey & Quigley, 2002). The PAI has demonstrated significant concurrent validity with measures used in forensic settings and correctional settings (Douglas et al., 2001; Duellman & Bowers, 2004).

MSI-2 measures. The original MSI (Nichols & Molinder, 1984) and the second version of the MSI, the MSI-2 (Nichols & Molinder, 2000), are theory-based self-report questionnaires used with sexual offenders designed to assess a wide range of psychosexual characteristics. The MSI-2 is an expanded version of the original 300-item MSI (Nichols & Molinder, 1984) and includes 560 true or false items that make up 40 scales and indices, including sexual dysfunction, psychosexual scales, accountability scales, behavioral scales, and several social desirability scales. Internal consistency estimates (majority of scales had Cronbach's α s of $.73$ to $.90$) and temporal reliability of the MSI-2 are good (Nichols & Molinder, 2000). Correlations of the MSI-2 scales with 13 scales of the Minnesota Multiphasic Personality Inventory-2 (MMPI-2) were small and

indicate that the MSI-2 is measuring many different aspects of sexuality, rather than one large construct (Nichols & Molinder, 2000). Cross validation studies with 12 independent samples of child molesters from across the United States showed no significant differences among the samples on 36 scales of the MSI-2. Among eight matched samples with known paraphilias, significant mean differences were found between samples on relevant MSI-2 scales except for the Masochism scale, indicating the ability of the MSI measure to differentiate between deviant sexual interests (Nichols & Molinder, 2000). More recently, Ennis, Buro, and Jung (2015) found differences on some of the MSI scales among low, low-moderate, and moderate-high risk groups of sex offenders.

Clarke Sex History Questionnaire-Revised (SHQ-R). The SHQ-R (Langevin & Paitich, 2006) is a self-report measure with 508 items that examine a wide range of conventional and deviant sexual behaviors. In total, 23 scales help to provide a comprehensive sexual history, including scales indicating engagement in deviant behaviors (e.g., exhibitionism, voyeurism, sexual abuse, fetishism). Although the SHQ-R is a published measure of self-reported participation in sexually deviant behaviors, no published studies have revealed the construct and predictive validity of the measure in its use with sexual offenders.

Cormier-Lang Criminal History Score. The Cormier-Lang scores are used to quantify an offender's history of criminal offenses, a current or index offense, or a particular subgroup of offenses and can be used when official police information is available (Quinsey et al., 2006). Different point values are assigned to different types of offenses (e.g., robbery = 7 points), and the score is tallied based on an offender's prior charges and convictions. Three types of scores were calculated for the present study; these include nonviolent, violent, and sexual offense scores (the latter score was specifically created for this study and includes violent sexual offenses only).

Procedures

Clinical files were reviewed at an outpatient forensic clinic to identify patients with a sexual offense on their records. The case files that were reviewed contained assessment reports, criminal records, case notes, offenders' reports, demographic information, and some description of victim information. Files were coded retrospectively, and no additional measures were administered for the purpose of this research. Because not all variables were available in the files for all offenders, sample sizes varied depending on the analysis. The present study is part of a larger database in which 406 variables were coded from the case files on each offender. To ensure that we maintained strong inter-rater reliability, research assistants received a full day of training on the variables. In addition, a small subset of case files were coded independently by two raters, and inter-rater reliability (reported elsewhere, Jung, Ennis, Stein, Choy, & Hook, 2013) was established for some of the risk measures and items and the offender characteristics variables (whereas some risk measures and items and psychometric scales were recorded directly from the case files and inter-rater reliability was not considered a concern). Two research assistants also rated 29 clinical files of sexual

Table 1. Descriptive Statistics, Internal Consistency, and Inter-Rater Reliability for the Static-2002R Items, Subscales, and Total Score.

Item and subscales	M (SD)	Range	Kappa or α^a	% ^b
1 Age at release	0.76 (1.3)	-2-2	.94	96.7
Persistence of Sex Offending subscale (2-4)	0.40 (0.78)	0-3	.58	
2 Sentencing occasions for sex offending	0.33 (0.69)	0-3	—	90.0
3 Juvenile arrest for sex offense	0.05 (0.23)	0-1	1.0	100
4 Rate of sex offending	0.11 (0.31)	0-1	.35	90
Deviant Sexual Interests Subscale (5-7)	0.42 (0.70)	0-3	.25	
5 Any non-contact sex offenses sentencing	0.19 (0.39)	0-1	.87	96.7
6 Any male victims	0.12 (0.33)	0-1	1.000	100
7 Young, unrelated victims	0.11 (0.32)	0-1	.44	86.7
Relationship to Victim Subscale (Items 8-9)	1.00 (0.73)	0-2	.53	
8 Unrelated victim	0.73 (0.45)	0-1	.71	86.7
9 Any stranger victims	0.27 (0.44)	0-1	.78	96.7
General Criminality Subscale (Items 10-14)	1.01 (0.97)	0-3	.79	
10 Prior criminal justice involvement	0.62 (0.49)	0-1	.93	96.7
11 Prior sentencing occasions	0.33 (0.54)	0-2	.92	96.7
12 Any past community supervision violation	0.23 (0.42)	0-1	.84	96.7
13 Years free prior to index offense	0.17 (0.38)	0-1	.43	86.7
14 Prior non-sexual violence sentencing	0.29 (0.45)	0-1	.49	76.7
Static-2002R total score	3.6 (2.59)	-2-12	ICC = .89	

Note. Bold values indicate alpha values. ICC = intra-class correlation.

^aKappa values for each item examine inter-rater reliability. Alpha coefficients examine each internal consistency.

^bPercentage agreement between raters is shown.

offenders by reviewing collateral documentation independently to examine inter-rater reliability, and these data are reported in the Results section.

Results

Descriptive statistics of the Static-2002R were calculated for the entire sample, and the means and standard deviations for items, content area subscales, and total scale score are presented in Table 1. Notably, our sample did not differ from the normative group upon which the Static-2002R and its predecessor were originally developed, as the average Static-2002R scores for both samples were not remarkably different ($M = 3.6$ in Phenix, Helmus, & Hanson, 2012). The following sections provide analyses that examine the reliability, concurrent validity, and factor structure of the Static-2002R.

Reliability of the Static-2002R

Examination of the internal properties of the Static-2002R led to calculations of Cronbach's alphas for the four subscales (see Table 1). The internal consistency of the

Table 2. Correlations Between Static-2002R Total and Subscales and the Static-99, SORAG, and SVR-20.

Instrument	Static-2002R total score	Subscales of the Static-2002R				
		Age	PSO	DSI	RV	GC
Static-2002R subscales						
Age	.44***		-.02	-.03	.14**	-.02
PSO	.49***			.37***	.23***	.44***
DSI	.41***				.26***	.21***
RV	.51***					.20***
GC	.53***					
Static-99R (<i>n</i> = 274-281)	.69***	.44***	.36***	.32***	.45***	.37***
SORAG (<i>n</i> = 112-114)	.48***	.41***	.22**	.02	.14	.44***
SVR-20 (<i>n</i> = 54)	.33***	.06	.28**	.24*	.12	.60***

Note. Kendall's tau correlations are reported. SORAG = Sex Offender Risk Appraisal Guide; SVR-20 = Sexual Violence Risk-20; PSO = Persistence of Sex Offending; DSI = Deviant Sexual Interests; RV = Relationship to Victim; GC = General Criminality.

* $p < .05$. ** $p < .01$. *** $p < .001$.

conceptualized subscales ranged from low to high (.25-.79)—only the GC subscale had good internal consistency, and the PSO subscale had moderate reliability. Interrater reliability of the Static-2002R was also examined. Kappa statistics, percentage agreements, and ICC were used, and these are listed in Table 1. When examining the kappa values (median = .84), 13 items had score variability that allowed for calculation of kappa. Items 1, 3, 5, 6, 8, 9, 10, 11, and 12 had values above 0.70, but Items 4, 7, 13, and 14 were below .50. In terms of percentage agreement, complete agreement (i.e., same rating given by each rater) was very good (median = 96.7%). The ICC was .89 (consistency) for the total Static-2002R score. Correlations between the content area subscales and the total score on the Static-2002R were also calculated, and all correlations were moderate to large and significant (see Table 2), with the exception of the age at release item. The inter-correlations among the four multi-item subscales were small to moderate and significant (r s from .20 to .44). The age item was moderately correlated with the total Static-2002R score ($r = .44$, $p < .001$) and had a small relationship with the Relationship to Victims subscale ($r = .14$, $p < .01$), but did not correlate with the other three subscales.

Concurrent Validity

Concurrent validity was examined by calculating correlation coefficients between the Static-2002R and three established sexual recidivism risk assessment instruments: Static-99R, SORAG, and SVR-20. Non-parametric statistical procedures were used (Gibbons & Chakraborti, 2003); specifically, Kendall's tau correlations between the Static-2002R total and content area subscale scores, and these criterion instruments

were calculated and are presented in Table 2. All correlations between the total score on the Static-2002R and the criterion instruments were significant. The largest observed correlation was with the Static-99R, $r(272) = .69, p < .001$. Significant relationships were also observed between the Static-2002R and the SORAG, $r(110) = .48, p < .001$, and the SVR-20 total score, $r(54) = .33, p < .01$. Varying results emerged when examining the associations between subscales and the criterion instruments. All subscales were significantly correlated to the Static-99R total score. However, the SORAG was positively correlated with the age at release item, and the PSO and GC subscales. Also, the SVR-20 total score was positively correlated with the PSO, DSI, and GC subscales but not with the age at release item or the Relationship to Victim subscale.

Construct Validity

Construct validity was explored by correlating three of the Static-2002R content area subscale scores (i.e., PSO, DSI, GC) with theoretically relevant constructs. In light of the positively skewed distributions of all three Static-2002R subscales, Kendall's tau non-parametric correlations were calculated.

PSO subscale. Kendall's tau correlation coefficients were calculated between the PSO subscale and selected items and scales from the SVR-20, the MSI measures, SHQ-R scales, the CATS, and the PAI, and the Cormier–Lang indices of criminal history. Tau correlation coefficients between the PSO subscale scores and relevant construct criteria variables are presented in Table 3.

Significant correlations were evident between the PSO subscale and variables associated with early onset of sexual and non-sexual behavioral problems, including the CATS total score ($r = .24$) and age at first criminal conviction ($r = -.20$). Notably, the PSO subscale was significantly associated with the Sexual Obsessions scale on the MSI-2 ($r = .34$). One variable directly associated with pedophilia demonstrated significant correlations with the PSO subscale; the SSPI had a moderate sized association ($r = .37$). Measures relevant to antisociality were also examined, and the PSO subscale was associated with the Antisocial scale of the PAI ($r = .22$), and the Cormier–Lang score for non-violent ($r = .30$), violent ($r = .41$), and sexually violent offenses ($r = .68$). Only one of the SVR-20 items, high density offenses, was associated with the PSO subscale ($r = .28$). The cognitive distortion variable and pornography index from the MSI-2, three selected items from the SVR-20 that measure frequency and diversity of sexual offending and psychopathy, and child sexual contact scales of the SHQ-R were not significantly associated with the PSO subscale.

DSI subscale. Associations between the DSI subscale and variables deemed relevant to the construct of sexual deviancy were examined, and Kendall's tau correlation coefficients are listed in Table 4. Criterion variables included the SSPI, as well as selected items and scales from the SVR-20, the MSI measure, and the SHQ-R.

The two variables most clearly representative of the DSI construct are the SSPI and the sexual deviance item on the SVR-20. The DSI subscale showed significant and

Table 3. Correlations Between the PSO Subscale of the Static-2002R and Constructs Reflecting Chronicity of Offending.

Construct	PSO	<i>n</i>
SVR-20		
Psychopathy	.12	60
High density offenses	.28*	65
Multiple offense types	.21	66
MSI-2		
Cognitive distortions/immaturity index	-.02	66
Sexual Obsessions scale	.34**	66
Pornography index	.23	66
SHQ-R		
Female child sex contact	.04	57
Male child sex contact	-.30	57
PAI		
Antisocial	.22***	163
Other relevant variables		
CATS	.24***	154
SSPI	.37***	329
Cormier–Lang non-violent offense score	.30***	328
Cormier–Lang violent offense score	.41***	330
Cormier–Lang sexual violence offense score	.68***	329
Age at first criminal conviction	-.20***	278

Note. Kendall's tau correlations are reported. PSO = Persistence of Sex Offending; SVR-20 = Sexual Violence Risk-20; MSI-2 = Multiphasic Sex Inventory; SHQ-R = Clarke Sex History Questionnaire-Revised; PAI = Personality Assessment Inventory; CATS = Childhood and Adolescent Taxon Scale; SSPI = Screening Scale for Pedophilic Interests.

* $p < .05$. ** $p < .01$. *** $p < .001$.

positive moderate correlations with both the SSPI ($r = .45$) and the SVR-20 Sexual Deviance item ($r = .37$), although the former should not be surprising given the item overlap between the DSI subscale items and the SSPI. A significant correlation between the DSI subscale and the child and adolescent sex experience scale of the SHQ-R emerged ($r = .41$), but emotional identification with children was not associated with the DSI subscale ($r = -.08$).

The DSI subscale was related to a few psychometric self-report variables reflecting paraphilic interests (namely, SHQ-R's exhibitionism and toucheurism/frotteurism), but it was unrelated to many paraphilic scales, including pedophilia-related scales on the MSI-2 and SHQ-R, and the pornography index. A significant association emerged between the DSI subscale and fantasy with males scale on the SHQ-R ($r = .32$). The Cormier–Lang sexual violence history score demonstrated moderate sized associations with the DSI subscale ($r = .29$). Self-reported frequency of past sexual contacts with children was not associated with DSI scores.

Table 4. Correlations Between the DSI Subscale of the Static-2002R and Constructs Reflecting Sexual Deviance.

Construct	DSI	<i>n</i>
SVR-20		
Sexual deviance	.37***	66
MSI-2		
Child molestation pattern	.13	66
Exhibitionist pattern	.16	66
Pornography index	.06	66
Fetishism and paraphilias NOS	.06	66
Bondage and discipline	.19	66
Sadism	-.03	66
Masochism	.05	66
SHQ-R		
Child and adolescent sex experience	.41***	56
Female child sex contact	-.04	56
Male child sex contact	.18	56
Child identification	-.08	55
Fantasy with males	.32*	56
Voyeurism	.21	56
Exhibitionism	.57***	56
Toucheurism and frotteurism	.23*	56
Other variables		
SSPI	.45***	328
Cormier–Lang sexual violence offense score	.29***	329

Note. Kendall's tau correlations are reported. DSI = Deviant Sexual Interests; SVR-20 = Sexual Violence Risk–20; MSI-2 = Multiphasic Sex Inventory; NOS = Not Otherwise Specified; SHQ-R = Clarke Sex History Questionnaire–Revised; SSPI = Screening Scale for Pedophilic Interests.

* $p < .05$. ** $p < .01$. *** $p < .001$.

GC subscale. Correlation coefficients were calculated between the GC subscale and selected items and scales from the SVR-20, the MSI measures, the SHQ-R, and the PAI, as well as the CATS. Relationships between the GC subscale and additional variables reflecting criminal history were also analyzed. Kendall's tau correlation coefficients are presented in Table 5.

The GC scale of the Static-2002R was significantly associated with numerous variables representing general antisociality. Regarding antisocial behaviors, such as criminality, large positive correlations were observed between the GC subscale with SVR-20 ratings of past non-sexual violence ($r = .53$), prior non-violent offending ($r = .64$), and also the Cormier–Lang offense history scores: non-violent offense history ($r = .65$), violent offense history ($r = .49$), and sexual violence offense history ($r = .31$). As expected, the rape pattern scale of the MSI-2 ($r = .25$) was significantly and positively correlated with the GC subscale, but the GC subscale did not reflect either the frequency of sex with

Table 5. Correlations Between the GC Subscale of the Static-2002R and Constructs Reflecting Non-Sexual Criminality.

Construct	GC	<i>n</i>
SVR-20		
Psychopathy	.36**	60
Substance abuse	.45***	66
Employment problems	.38***	68
Past non-sexual violence	.53***	68
Past non-violent offenses	.64***	68
Past supervision failure	.55***	67
Negative treatment attitudes	.10	65
MSI-2		
Rape pattern	.25*	66
Conduct disorder index	.38***	66
Sociopathy index	.52***	66
Antisocial behavior scale	.46***	66
Substance abuse index	.30**	66
SHQ-R		
Female adult sex frequency	.05	57
Sexual aggression	.20	56
PAI		
Antisocial	.23***	163
Antisocial behavior	.30***	160
Antisocial egocentricity	.11	160
Antisocial stimulation seeking	.11	160
Alcohol	.31***	163
Drug	.20***	163
Aggression	.19**	163
Treatment rejection	-.06	163
Other relevant variables		
CATS	.35***	154
Cormier–Lang non-violent offense score	.65***	329
Cormier–Lang violent offense score	.49***	332
Cormier–Lang sexual violence offense score	.31***	331
Age at first criminal conviction	-.41***	279

Note. Kendall's tau correlations are reported. GC = General Criminality; SVR-20 = Sexual Violence Risk-20; MSI-2 = Multiphasic Sex Inventory; SHQ-R = Clarke Sex History Questionnaire-Revised; PAI = Personality Assessment Inventory; CATS = Childhood and Adolescent Taxon Scale.

* $p < .05$. ** $p < .01$. *** $p < .001$.

adult females or scores on the sexual aggression scale, as measured on the SHQ-R. Age at first criminal conviction demonstrated a significant negative correlation with the GC Subscale ($r = -.41$), indicating that the younger the offender, the more likely they had a higher score on general criminality.

The GC subscale was significantly moderately associated with numerous psychometric measures of criminality and antisociality (MSI-2: sociopathy index, $r = .52$, conduct disorder index, $r = .38$, and antisocial behavior scales, $r = .46$; CATS total score, $r = .35$; PAI: antisocial scale, $r = .23$, and antisocial behavior subscale, $r = .30$) and the psychopathy item of the SVR-20 ($r = .36$).

The GC subscale was also significantly positively related to a range of variables peripherally associated with criminality, including multiple indicators of substance abuse (i.e., SVR-20, MSI-2, PAI; all $ps < .001$), employment problems (i.e., SVR-20), past supervision failure (i.e., SVR-20), and aggression (i.e., PAI). Treatment rejection, as measured on both the SVR-20 and PAI, were not correlated with the GC subscale.

Factor Structure

Exploratory factor analysis (EFA) was used to investigate the factor structure of the Static-2002R. The EFA analyses were computed using MPlus version 7.3 (Muthén & Muthén, 2014). When conducting an EFA, there are several important methodological considerations that can impact the appropriateness and generalizability of the results (see Schmitt, 2011, for an overview); these include the type of correlation matrix analyzed, the method used to estimate the model, the rotation method, and the method used to determine the number of factors and model fit. With these considerations in mind, the following methodology was used in the current study: Factors were extracted from a polychoric correlation matrix using the Weighted Least Square (WLSMV) estimation method, and factors were rotated using oblique rotation (Geomin).

Participants with missing data were excluded from the EFA ($n = 30$) as MPlus does not allow for missing data when factor analyzing categorical data (Muthén & Muthén, 2014); this resulted in a final sample size of 342. Four factor retention methods were examined to determine the number of factors to retain in the model: (a) Kaiser's Criterion (i.e., keeping the number of factors with eigenvalues greater than 1) suggested four factors, (b) Scree Plot suggested two to four factors, (c) Parallel Analysis (O'Connor, 2000) suggested three factors, and (d) Velicer's Minimum Average Partial test (MAP test; O'Connor, 2000) suggested two factors. Models with two to four factors were computed and fit indices were used to determine overall factor structure fit. Items were considered to load on a factor if the factor loading was at least .40 (Cudeck & O'Dell, 1994; Schmitt & Sass, 2011) and the factor loading significantly differed from 0 (Cudeck & O'Dell, 1994; Schmitt, 2011; for a more detailed discussion of the EFA methodology used, see Hermann, Babchishin, Nunes, Leth-Steensen, & Cortoni, 2012).

EFA models with three or four factors that included all 14 Static-2002R items were problematic as they resulted in one or more items having negative residual variance (i.e., model explains more variance in an item than is observed). The two factor model, however, demonstrated acceptable fit with a root mean square error of approximation (RMSEA) of .059 (90% confidence interval [CI] = [.046, .072]), comparative fit index (CFI) of .97, and standardized root mean square residual (SRMR) of .116, and did not

Table 6. Rotated Factor Loadings for Exploratory Factor Analysis of Static-2002R Items.

Static-2002R item	Rotated factor loadings (standardized score)	
	Factor 1	Factor 2
Age at release	-.105 (-1.095)	.113 (1.266)
Prior sentencing for sex offending	.725^a (10.009)	.262 (2.407)
Juvenile arrest for sex offending	.355 (2.097)	.342 (2.438)
High rate of sex offending	.855^a (9.547)	.253 (1.744)
Any sentencing for non-contact sex offending	.737^a (7.753)	-.165 (-1.247)
Any male victims	.360 (3.184)	.033 (0.280)
Young, unrelated victim	.676^a (8.723)	.007 (0.191)
Any unrelated victims	.488^a (4.621)	-.011 (-0.114)
Any stranger victims	.671^a (7.401)	-.111 (-0.885)
Any prior criminal justice involvement	.289 (3.707)	.795^a (11.183)
Any prior sentencing occasions	-.003 (-0.399)	.794^a (36.201)
Any community supervision violations	.026 (0.308)	.802^a (12.691)
Years free prior to index	.279 (2.804)	.652^a (7.008)
Prior non-sexual violence sentencing	-.270 (-2.589)	.965^a (13.001)

Note. For the standardized score, the critical z score was 3.13, $\alpha = .0009$. Bolded values indicate significant loading on the factor ($p < .0009$; Cudeck & O'Dell, 1994).

^aLoading above .40.

have items with negative residual variance. As a result, the two factor model was selected for interpretation. The rotated factor loadings and their associated standardized scores are presented in Table 6.

The first factor consisted of items assessing PSO (Items 2, 4), DSI (Items 5, 6, 7), and Relationship to Victims (Items 8, 9). The second factor consisted of the items assessing GC (Items 10, 11, 12, 13, 14). Age did not load onto either of the factors. The item assessing juvenile arrests for sexual offending did not load significantly on to factors 1 or 2 (but the factor loadings were not trivial); conceptually, the item fits with the first factor, and as a result was included in factor 1. The internal consistency was low for factor 1 ($\alpha = .66$) and good for factor 2 ($\alpha = .81$). As expected, the two factors have a moderate to large positive relationship ($r = .456, p < .05$).

Discussion

The present study provides an important and necessary independent examination of the reliability, construct validity, and factor structure for three of the content areas measured by the Static-2002R. The subscales were compared with the items and scales of both extant risk assessment measures and several psychometric non-risk measures. Our overall findings provide moderate support for the reliability and strong support for the construct validity of the Static-2002R subscales.

Reliability was not always consistent across the items on the Static-2002R, its subscales, or across raters. Of the 4 multi-item subscales, only the GC subscale demonstrated good internal consistency, whereas the other three subscales had low to moderate internal consistency. Our findings are consistent with other research that has examined the internal consistency of the Static-2002 (i.e., GC had the highest reported alphas, with DSI and Relationship to Victim subscales having the lowest in Langton, Barbaree, Hansen, et al., 2007). Most of the items were reliably scored with the exception of 4 items (i.e., rate of sex offending; young, unrelated victims; years free prior to index offense; prior non-sexual violence sentencing). These 4 items were not consistently classified by the two raters, but percentage agreement was within acceptable limits. Notably, 4 of the 14 Static-2002R items demonstrated relatively poor inter-rater reliability. For the first 3 of those items (i.e., Item 4, Rate of sex offending; Item 7, Young, unrelated victims; and Item 13, Years free prior to index offense), poor inter-rater reliability may be attributable to shortcomings in the training of the research assistants in combination with the relative complexity of the items' rating criteria. Unlike the majority of items on the Static-2002R, these items have multifaceted contingencies for coding. For example, the criteria for assigning a score for Item 7, Young, Unrelated Victims, is actually threefold, and requires the rater to consider the number of victims, the age of those victims, and the nature of the relationship between perpetrator and victim. For the purpose of this study, Static-2002R ratings were completed by research assistants who were trained by the primary researchers, rather than by a certified Static-2002R trainer; moreover, not all research assistants were trained at the same time, or by the same trainer, and, therefore, there may have been discrepancies and/or deficiencies in certain aspects of training between assistants. It is reasonable to speculate that deficiencies in training would be most evident in relation to more complex Static-2002R items, such as these.

Item 14, Any Prior Non-sexual Violence Sentencing Occasions, also demonstrated poor inter-rater reliability, but unlike Items 7 and 13, this item does not require consideration of multiple criteria to arrive at a rating of 0 or 1. For this item, inconsistent ratings may be attributable to the fact that, in most cases, research assistants would not have had direct access to each subjects' official criminal record, but rather, would have had to rely on file information, clinical case notes, and the final report of the assessing clinician. The quality and clarity of historical information would be expected to vary from subject to subject, and this variability could negatively impact inter-rater reliability. Past reports have shown good reliability between raters, so perhaps the use of mental health files may be more challenging to consistently and confidently score these particular items, given the multiple and varied sources of information in each file.

Concurrent validity of the Static-2002R with other measures of sexual violence risk (i.e., Static-99R, SORAG, and the SVR-20) was examined. It is not surprising that the Static-99R was consistent with the total score on the Static-2002R and its five subscales. Consistent with other reports (e.g., Langton, Barbaree, Seto, et al., 2007; Parent et al., 2011), the Static-2002R and three of its subscales were correlated with the total score on the SORAG and the SVR-20. Both of these criterion measures were

correlated to the PSO and GC subscales. The SORAG was also associated with the age item, which was not surprising given the SORAG contains an age variable among its 14 items. Similar to Langton, Barbaree, Seto, et al.'s (2007) finding, our data failed to yield a significant correlation between the DSI subscale of the Static-2002R and the SORAG total score. The absence of strong relationships between SORAG scores and the DSI subscale may be explained by the fact that, despite being a valid predictor of violent recidivism by sex offenders, 11 of 14 SORAG items are factors associated with non-sexual violence and criminality. In contrast, the DSI items relate specifically to sexual behavior or interests. The SVR-20 was associated with DSI, which is consistent with the inclusion of construct-related items in the SVR-20 (e.g., sexual deviance, diversity of sexual violence). Hence, the Static-2002R and its subscales demonstrate strong concurrent validity in our analyses.

What was of unique interest was the association between three of the subscales of the Static-2002R and related constructs on non-risk assessment measures, namely psychometric self-report inventories. A fourth subscale, the Relationship to Victims subscale, comprised two items, and its validity was not examined; however, it demonstrated good inter-rater reliability.

The PSO subscale demonstrated several linkages with related constructs that included early onset of sexual and non-sexual behavioral difficulties and a psychometric scale of sexual obsession, as well as screening measures of pedophilia and chronic antisociality, and, therefore, construct validity was supported. The association between the DSI subscale and selected convergent measures of deviancy indicated that the subscale had good construct validity. The subscale was strongly associated with variables deemed most construct-relevant (i.e., SSPI and sexual deviance item on the SVR-20) and assessed interests rather than behaviors (i.e., sexual contact with children). Although the subscale was not associated with all forms of paraphilic interests, there were demonstrated relationships with some paraphilias, including toucheurism and exhibiting behavior, which corresponds to what other researchers have posited (i.e., those with paraphilias are often diversified in their interests; Abel, Becker, Cunningham-Rathner, Mittelman, & Rouleau, 1988; Laws & O'Donohue, 2008).

In examining the GC subscale, support for its construct validity was strong. Nearly all the variables included in the analyses were associated with scores on this subscale, reflecting antisociality, past non-sexual offending, violent behavior, psychometric measures of antisociality and conduct disorder, and substance abuse scales. In addition to the subscale's strong internal consistency and its demonstrated convergent validity, the subscale's items loaded on a single factor, demonstrating a commonality of its items and their association to a single construct.

The EFA revealed a two factor model. The first factor primarily consisted of items related to sexual offending specifically (persistence in sexual offending, deviant sexual interests, and relationship to victim) whereas the second factor consisted of the GC subscale. Although this is not consistent with Langton, Barbaree, Hansen, et al.'s (2007) five-factor solution for the Static-2002, our findings are in line with the factor analytic findings for the Static-99. Several studies have consistently found three-factor models of sexual deviancy and criminality, along with a third less-defined factor of

detachment, using the items from the Static-99R (e.g., Allen & Pflugradt, 2014; Brouillette-Alarie & Proulx, 2013). In contrast to Langton, Barbaree, Hansen, et al. (2007), the current study used EFA methodology that is optimal for ordinal data (i.e., polychoric correlations, WLSMV estimator; see Schmitt, 2011, for an overview). This is one possible explanation for the differences in factor structure across studies.

As stated earlier, there is often an emphasis on the importance of demonstrating a risk assessment measure's predictive validity. However, the extent to which a measure is capturing the constructs of interest is also an important issue and has implications for the discordance seen among sexual violence risk measures (see Babchishin et al., 2012, for further discussion). Clinicians and researchers alike should not take for granted that the measure is adequately assessing the intended constructs, but rather should analyze the measure's association with variables that are external to the measure (AERA, APA, & NCME, 1999). As noted by Babchishin et al. (2012), identifying latent constructs of interest and using multifaceted evidence to examine whether the measure is assessing these latent constructs is a critical first step in establishing construct validity. Understanding these latent constructs has implications for the discordance among measures as seen in many studies (e.g., Barbaree, Langton, & Peacock, 2006; Jung, Pham, & Ennis, 2013). With such evidence, one can feel greater confidence that the risk assessment tool is measuring certain domains. This is particularly true in the case of the Static-2002R, which was designed to include clinically relevant subscales that may inform clinical formulations, recommendations, and interventions. The findings in this examination of the Static-2002R provide moderate to strong support for the assumption that the subscales of the Static-2002R are measuring what they purport to measure. The concepts that are being assessed by the three subscales, PSO, DSI, and GC, appear to have convergence with related measures and discriminate from unrelated variables. Contrary to previous findings, the reliability of the Static-2002R, as summarized in this investigation, was less than what would be preferred and expected, and we can only conclude that there is a moderate level of consistency across raters and within the subscales. Although this conservative finding may be attributed to the sample collected in this particular study, it does raise some concern with regard to the homogeneity of the items within each subscale.

It is important to note several limitations of the present study. First, there were smaller sizes used in some analyses. Although we started with a fairly large sample of sex offenders, not all offenders had completed psychometric measures (e.g., many of the analyses that included the SHQ-R had correlational analyses based on sizes as small as 70). Regarding the lack of observed association between the GC subscale of the Static-2002R and the PCL-R item of the SVR-20, it must be noted that only rarely did the reviewed clinical files include formal psychopathy assessments using the PCL-R. Correspondingly, we must presume that in cases where the SVR-20 was included in the file, and where a rating was provided for the Psychopathy item, that item was likely scored based on an informal, unstructured clinical assessment by the clinician who completed the SVR-20. Consequently, the lack of an association between the GC scale of the Static-2002R and the Psychopathy item of the SVR-20 must be viewed with caution and consideration for the influence of questionable data.

The EFA analyses were also limited in that, due to the design of the Static-2002R, often there were not multiple items representing each of the risk-relevant constructs being assessed. Ideally, there would be at least four to five items representing each latent construct in an EFA. This is problematic as it is possible Factor 1 may have separated into multiple factors had there been more items representing each of the purportedly underlying risk-relevant constructs being assessed (e.g., persistence in sex offending, deviant sexual interests). In support of this notion, the internal consistency of factor 1 was low ($\alpha = .66$), suggesting the items are heterogeneous. Future research can address this issue by conducting EFA analyses with the Static-2002R items and additional indicators of each purportedly assessed risk-relevant construct. If the items from Static-2002R load onto a factor with the additional items included representing its associated construct, then that would provide evidence of construct validity for the subscale items of Static-2002R.

Finally, our results indicated that, although most items on the Static-2002R were reliably scored, four items (i.e., Rate of sex offending; Young, unrelated victims; Years free prior to index offense; Prior non-sexual violence sentencing) demonstrated substantially poorer reliabilities. These four items were not consistently classified by the two raters, but percentage agreement was within acceptable limits. It is worth noting that three of the four items that demonstrated poor inter-rater reliability were among the more complex of the Static-2002R items to code in that they involve scoring criteria with multiple contingencies. This suggests that more intensive training in Static-2002R coding for research assistants might have improved inter-rater reliability for these more complex items.

In conclusion, this study makes a relevant contribution to the risk prediction literature. Although it might be argued that examination of a risk appraisal instrument's psychometric properties is less stimulating than studies investigating the predictive capabilities of such measures, this view does not negate the need for such research. In addition to the plethora of evidence supporting the Static-2002R as a risk prediction tool, our findings advocate for the use of the Static-2002R as a measure of factors that are deemed relevant in predicting risk, as corroborated by their associations with construct-related variables. In light of the frequency with which the Static-2002R is used in correctional, criminal justice, and outpatient forensic settings, it is crucial that the measure is continually re-examined and revised accordingly to ensure that it remains a psychometrically sound measure.

Acknowledgments

The authors gratefully acknowledge the support provided by the Northern Alberta Forensic Psychiatry Services, the Faculty of Arts & Science and the Research Office of MacEwan University, and the Integrated Threat and Risk Assessment Centre of the Alberta Law Enforcement Response Teams. They also thank the research assistance of Melissa Daniels, John DeCesare, Barbara Ann Dowker, Lisa Jamieson, Bazyl Stefanski, and Shayla Stein, and the valuable feedback of anonymous reviewers.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

- Abel, G. G., Becker, J. V., Cunningham-Rathner, J., Mittelman, M., & Rouleau, J. L. (1988). Multiple paraphilic diagnoses among sex offenders. *Bulletin of the American Academy of Psychiatry and the Law, 16*, 153-168.
- Allen, B. P., & Pflugradt, D. M. (2014). An exploration of the latent constructs of the Static-99. *International Journal of Offender Therapy and Comparative Criminology, 58*, 1376-1388. doi:10.1177/0306624X13496046
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Archer, R. P., Buffington-Vollum, J. K., Stredny, R. V., & Handel, R. W. (2006). A survey of psychological test use patterns among forensic psychologists. *Journal of Personality Assessment, 87*, 84-94. doi:10.1207/s15327752jpa8701_07
- Babchishin, K. M., Hanson, R. K., & Helmus, L. (2012). Even highly correlated measures can add incrementally to actuarial risk prediction. *Assessment, 19*, 442-461. doi:10.1177/1073191112458312
- Barbaree, H. E., Langton, C. M., & Peacock, E. J. (2006). Different actuarial risk measures produce different risk rankings for sexual offenders. *Sexual Abuse: A Journal of Research and Treatment, 18*, 423-440. doi:10.1177/107906320601800408
- 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*, 490-521. doi:10.1177/009385480102800406
- Bartosh, D. L., 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*, 422-438. doi:10.1177/0306624X03253850
- Bengtson, S. (2008). Is newer better? A cross-validation of the Static-2002 and the Risk Matrix 2000 in a Danish sample of sexual offenders. *Psychology, Crime & Law, 14*, 85-106. doi:10.1080/10683160701483104
- Bengtson, S., & Langstrom, N. (2007). Unguided clinical and actuarial assessment of re-offending risk: A direct comparison with sex offenders in Denmark. *Sexual Abuse: A Journal of Research and Treatment, 19*, 135-153. doi:10.1007/s11194-007-9044-5
- Boer, D. P., Hart, S. D., Kropp, P. R., & Webster, C. D. (1997). *Manual for the Sexual Violence Risk-20: Professional guidelines for assessing risk of sexual violence*. Vancouver, Canada: The Mental Health, Law, & Policy Institute.
- Boone, D. (1998). Internal consistency reliability of the Personality Assessment Inventory with psychiatric inpatients. *Journal of Clinical Psychology, 54*, 839-843. doi:10.1002/(SICI)1097-4679(199810)54:6<839::AID-JCLP12>3.0.CO;2-M
- Boyle, G. J., & Lennon, T. J. (1994). Examination of the reliability and validity of the Personality Assessment Inventory. *Journal of Psychopathology and Behavior Assessment, 16*, 173-188. doi:10.1007/BF02229206
- Brouillette-Alarie, S., & Proulx, J. (2013). Predictive validity of the Static-99R and its dimensions. *Journal of Sexual Aggression, 19*, 311-328. doi:10.1080/13552600.2012.747630

- Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. *Psychological Bulletin, 115*, 475-487. doi:10.1037/0033-2909.115.3.475
- de Vogel, V., de Ruiter, C., van Beek, D., & Mead, G. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and Human Behavior, 28*, 235-251. doi:10.1023/B:LAHU.0000029137.41974.eb
- Douglas, K. S., Hart, S. D., & Kropp, P. R. (2001). Validity of the Personality Assessment Inventory for forensic assessments. *International Journal of Offender Therapy and Comparative Criminology, 45*, 183-197. doi:10.1177/0306624X01452005
- Duellman, R. M., & Bowers, T. G. (2004). Use of the Personality Assessment Inventory (PAI) in forensic and correctional settings: Evidence for concurrent validity. *International Journal of Forensic Psychology, 1*, 42-57.
- Ennis, L., Buro, K., & Jung, S. (2015). Identifying male sexual offender subtypes using cluster analysis. *Sexual Abuse: A Journal of Research and Treatment*. Advance online publication. doi:10.1177/1079063214527481
- Gibbons, J. D., & Chakraborti, S. (2003). *Nonparametric statistical inference (4th ed.)*. New York, NY: Marcel Dekker.
- Glover, A. J. J., Nicholson, D. E., Hemmati, T., Bernfeld, G. A., & Quinsey, V. L. (2002). A comparison of predictors of general and violent recidivism among high-risk federal offenders. *Criminal Justice and Behavior, 29*, 235-249. doi:10.1177/0093854802029003001
- Hanson, R. K. (2009). The psychological assessment of risk for crime and violence. *Canadian Psychology, 50*, 172-182. doi:10.1037/a0015726
- Hanson, R. K., Helmus, L., & Thornton, D. (2010). Predicting recidivism among sexual offenders: A multi-site study of Static-2002. *Law and Human Behavior, 34*, 198-211. doi:10.1037/t04165-000
- Hanson, R. K., Lunetta, A., Phenix, A., Neeley, J., & Epperson, D. (2014). The field validity of the Static-99/R sex offender risk assessment tool in California. *Journal of Threat Assessment and Management, 1*, 102-117. doi:10.1037/tam0000014
- Hanson, R. K., & Morton-Bourgon, K. (2009). The accuracy of recidivism risk assessment for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment, 21*, 1-21. doi:10.1037/a0014421
- Hanson, R. K., & Thornton, D. (1999). *Static-99: Improving actuarial risk assessments for sex offenders* (User Report No. 1999-02). Ottawa, Canada: Solicitor General of Canada.
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of the Static-2002* (User Report No. 2003-01). Ottawa, Canada: Solicitor General Canada.
- Harris, A., Phenix, A., Hanson, R. K., & Thornton, D. (2003). *Static-99 coding rules revised—2003*. Ottawa, Canada: Solicitor General of Canada.
- Harris, G. T., Rice, M. E., & Quinsey, V. L. (1994). Psychopathy as a taxon: Evidence that psychopaths are a discrete class. *Journal of Consulting and Clinical Psychology, 62*, 387-397. doi:10.1037/0022-006X.62.2.387
- Harris, G. T., Rice, M. E., Quinsey, V. L., Lalumière, M. L., Boer, D., & Lang, C. (2003). A multisite comparison of actuarial risk instruments for sex offenders. *Psychological Assessment, 15*, 413-425. doi:10.1037/1040-3590.15.3.413
- Helmus, L., & Hanson, K. (2007). Predictive validity of the Static-99 and Static-2002 for sex offenders on community supervision. *Sexual Offender Treatment, 2*, 1-14.
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment, 24*, 64-101. doi:10.1177/1079063211409951

- Hermann, C. A., Babchishin, K. M., Nunes, L. K., Leth-Steensen, C., & Cortoni, F. (2012). Factor structure of the Bumby RAPE Scale: A two-factor model. *Criminal Justice and Behavior, 39*, 869-886. doi:10.1177/0093854812436802
- Jung, S., Daniels, M. K., Friesen, M., & Ledi, D. (2012). An examination of convergent constructs among Level of Service measures and other measures. *Journal of Forensic Psychiatry & Psychology, 23*, 601-619. doi:10.1080/14789949.2012.732595
- Jung, S., Ennis, L., Stein, S., Choy, A. L., & Hook, T. (2013). Child pornography possessors: Comparisons and contrasts with contact- and non-contact sex offenders. *Journal of Sexual Aggression, 19*, 295-310. doi:10.1080/13552600.2012.741267
- Jung, S., Pham, A., & Ennis, L. (2013). Measuring the disparity of categorical risk among various sex offender risk assessment instruments. *Journal of Forensic Psychiatry & Psychology, 24*, 353-370. doi:10.1080/14789949.2013.806567
- Langevin, R., & Paitich, D. (2006). *The Clarke Sex History Questionnaire for Males-Revised (SHQ-R): Technical manual*. Toronto, Canada: Multi-Health Systems.
- Langton, C. M. (2003). *Contrasting approaches to risk assessment with adult male sexual offenders: An evaluation of recidivism prediction schemes and the utility of supplementary clinical information for enhancing predictive accuracy* (Unpublished doctoral dissertation). University of Toronto, Ontario, Canada.
- Langton, C. M., Barbaree, H. E., Hansen, K. T., Harkins, L., & Peacock, E. J. (2007). Reliability and validity of the Static-2002 among adult sexual offenders with reference to treatment status. *Criminal Justice and Behavior, 34*, 616-640. doi:10.1177/0093854806296851
- Langton, C. M., Barbaree, H. E., Seto, M. C., Peacock, E. J., Harkins, L., & Hansen, K. T. (2007). Actuarial assessment of risk for reoffense among adult sex offenders: Evaluating the predictive accuracy of the Static-2002 and five other instruments. *Criminal Justice and Behavior, 34*, 37-59. doi:10.1177/0093854806291157
- Laws, D. R., & O'Donohue, W. T. (Eds.). (2008). *Sexual deviance: Theory, assessment, and treatment*. New York, NY: Guilford Press.
- Looman, J., & Abracen, J. (2010). Comparison of measures of risk for recidivism in sexual offenders. *Journal of Interpersonal Violence, 25*, 791-807. doi:10.1177/0886260509336961
- Mann, R. E., Hanson, R. K., & Thornton, D. (2010). Assessing risk for sexual recidivism: Some proposals on the nature of psychologically meaningful risk factors. *Sexual Abuse: A Journal of Research and Treatment, 22*, 191-217. doi:10.1177/1079063210366039
- McGrath, R., Cumming, G., Burchard, B., Zeoli, S., & Ellerby, L. (2010). *Current practices and emerging trends in sexual abuser management: The Safer Society 2009 North American Survey*. Brandon, VT: Safer Society Press.
- Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Morey, L. C. (2007). *The Personality Assessment Inventory (PAI): Professional manual* (2nd ed.). Lutz, FL: Psychological Assessment Resources.
- Morey, L. C., & Quigley, B. D. (2002). The use of the Personality Assessment Inventory (PAI) in assessing offenders. *International Journal of Offender Therapy and Comparative Criminology, 46*, 333-349. doi:10.1177/03024X02046003007
- Murrie, D. C., Boccaccini, M. T., Turner, D. B., Meeks, M., Woods, C., & Tussey, C. (2009). Rater (dis)agreement on risk assessment measures in sexually violent predator proceedings: Evidence of adversarial allegiance in forensic evaluation? *Psychology, Public Policy, and Law, 15*, 19-53. doi:10.1037/a0014897
- Muthén, L. K., & Muthén, B. O. (2014). *MPlus user's guide*. Los Angeles, CA: Author.

- Nichols, H. R., & Molinder, I. (1984). *Manual for the Multiphasic Sex Inventory*. Tacoma, WA: Crime and Victim Psychology Specialists.
- Nichols, H. R., & Molinder, I. (2000). *MSI II: Multiphasic Sex Inventory II handbook*. Tacoma, WA: Author.
- Nunes, K. L., & Babchishin, K. M. (2011). Construct validity of Stable-2000 and Stable-2007. *Sexual Abuse: A Journal of Research and Treatment*, 24, 29-45. doi:10.1177/1079063211404921
- Nunes, K. L., Firestone, P., Bradford, J. M., Greenberg, D. M., & Broom, I. (2002). A comparison of modified versions of the Static-99 and the Sex Offender Risk Appraisal Guide (SORAG). *Sexual Abuse: A Journal of Research and Treatment*, 14, 253-269. doi:10.1037/t01167-000
- O'Connor, B. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instruments, and Computers*, 32, 396-402. doi:10.3758/BF03200807
- Parent, G., Guay, J. P., & Knight, R. A. (2011). An assessment of long-term risk of recidivism by adult sex offenders: One size doesn't fit all. *Criminal Justice and Behavior*, 38, 188-209. doi:10.1177/0093854810388238
- Phenix, A., Doren, D., Helmus, L., Hanson, R. K., & Thornton, D. (2008). *Coding rules for Static-2002*. Ottawa, Canada: Public Safety Canada.
- Phenix, A., Helmus, L., & Hanson, R. K. (2012). *Static-99R & Static-2002R: Evaluators' workbook*. Available from www.static99.org
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Quinsey, V. L., Rice, M. E., & Harris, G. T. (1995). Actuarial prediction of sexual recidivism. *Journal of Interpersonal Violence*, 10, 85-105. doi:10.1177/088626095010001006
- Rettenberger, M., Matthes, A., Boer, D. P., & Eher, R. (2010). Prospective actuarial risk assessment: A comparison of five risk assessment instruments in different sexual offender subtypes. *International Journal of Offender Therapy and Comparative Criminology*, 54, 169-186. doi:10.1177/0306624X08328755
- Schmitt, T. A. (2011). Current methodological considerations in exploratory and confirmatory factor analysis. *Journal of Psychoeducational Assessment*, 29, 304-321. doi:10.1177/0734282911406653
- Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: Implications for factor pattern loadings and interfactor correlations. *Educational and Psychological Measurement*, 71, 95-113. doi:10.1177/0013164410387348
- Seto, M. C., Harris, G. T., Rice, M. E., & Barbaree, H. E. (2004). The Screening Scale for Pedophilic Interests predicts recidivism among adult sex offenders with child victims. *Archives of Sexual Behavior*, 33, 455-466. doi:10.1023/B:ASEB.0000037426.55935.9c
- Seto, M. C., & Lalumière, M. L. (2001). A brief screening scale to identify pedophilic interests among child molesters. *Sexual Abuse: A Journal of Research and Treatment*, 13, 15-25. doi:10.1177/107906320101300103
- Seto, M. C., Murphy, W. D., Page, J., & Ennis, L. (2003). Detecting anomalous sexual interests among juvenile sex offenders. *Annals of the New York Academy of Sciences*, 989, 106-118. doi:10.1111/j.1749-6632.2003.tb07298.x
- Stalans, L. J., Hacker, R., & Talbot, M. E. (2010). Comparing nonviolent, other-violent, and domestic batterer sex offenders: Predictive accuracy of risk assessments on sexual recidivism. *Criminal Justice and Behavior*, 37, 613-628. doi:10.1037/t04165-000