

Assessing Dynamic Risk Factors in Institutional Settings Using STABLE-2007

Yolanda Fernandez¹

[1] *Forensic Behaviour Services, Kingston, Ontario, Canada.*

Sexual Offending: Theory, Research, and Prevention, 2021, Vol. 16, Article e4551, <https://doi.org/10.5964/sotrap.4551>

Received: 2020-10-13 • **Accepted:** 2021-06-10 • **Published (VoR):** 2021-12-23

Handling Editor: Mark E. Olver, University of Saskatchewan, Saskatoon, SK, Canada

Corresponding Author: Yolanda Fernandez, 303 Bagot St. Suite 306, Kingston, Ontario, K7K 5W7, Canada. E-mail: ymfernandez01@gmail.com

Related: This article is part of the SOTRAP Special Thematic Section “Dynamic Assessment of Men Facing Indeterminate Detention”, Guest Editors: Andrew J. R. Harris & Mark E. Olver, *Sexual Offending: Theory, Research, and Prevention*, 16, <https://doi.org/10.5964/sotrap.v16>

Abstract

Assessing dynamic risk factors for persons who reside in an institution can be a challenge. Conceptualizing and scoring dynamic risk factors is more difficult when environments are restricted and opportunities for those being assessed to demonstrate changes in behaviour may be few and far between. Additionally, because dynamic risk measures rely partly on file information scoring is dependent on the training and backgrounds of the people who record information and their personal decisions as to what they consider important enough to include in records. This may mean that scoring under research conditions based only on file review does not reflect the reliability of the measure under clinical conditions. Despite these challenges the present paper argues that there is sufficient evidence to support the use of STABLE-2007 as a reliable and valid measure of dynamic risk factors in institutional settings under both clinical and research conditions. Tips are provided on how to conceptualize institutional behaviours in a manner relevant to dynamic risk factors and how to weigh historical versus more recent information. Finally, recommendations are made for implementing a thoughtful system of checks and balances relevant to the assessment process in institutional settings.

Keywords

dynamic assessment, STABLE, reliability, validity, institutions



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License, CC BY 4.0](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction, provided the original work is properly cited.

Non-Technical Summary

Background

The clinical utility of a dynamic risk factor assessment tool such as STABLE-2007 lies in its focus on identifying relevant intervention targets, contributing to the determination of appropriate treatment intensity levels, assessing offender progress in addressing their treatment needs, and monitoring on-going risk potential risk levels. However, assessing dynamic risk factors can be challenging within an institutional setting.

Why was this study done?

Dynamic risk assessments completed in institutional environments pose a number of challenges. These limitations can foster legal arguments where the defence claims that there is “no evidence of current risk” and the prosecution claims there is “no evidence of change”. Evaluators feel less comfortable making definitive statements about change and clients feel frustrated and demoralized when their perception is they have worked hard to adjust and adapt their behaviours, and this is not reflected in risk assessments.

What did the researchers do and find?

The present paper reviews current studies evaluating the reliability and validity of the STABLE measures (STABLE-2000 and STABLE-2007) scored in institutional settings. The author concludes there is sufficient evidence to support the use of STABLE-2007 as a reliable and valid measure of dynamic risk factors under both clinical and research conditions.

What do these findings mean?

Suggestions are made for conceptualizing institutional behaviours in a manner that integrates current behaviours into the dynamic risk evaluation. Finally, the author recommends ways to ensure high quality, reliable, and valid assessments that can confidently be used in decision making.

Highlights

- The clinical utility of a dynamic risk factor assessment tool such as STABLE-2007 lies in its focus on identifying relevant intervention targets.
- Completing STABLE-2007 assessments in an institutional environment poses a number of challenges.
- While the existing data indicates STABLE-2007 can be reliably scored in an institutional, more research is needed specific to the interrater reliability of scoring for cases that have been incarcerated for lengthy periods (e.g., 10+ years).
- The available studies generally support the use of the STABLE-2007 in institutional settings as a reliable and clinically useful measure.
- Evaluators should pay attention to the conditions under which scoring is completed, how file information is interpreted, interviewing techniques, and diligent gathering of collateral information, all of which may impact the reliability and utility of the measure.

Individuals involved in the treatment, supervision, and management of men who have committed sexual offences value the information garnered from a comprehensive assessment of risk. The clinical utility of a dynamic risk factor assessment tool such as STABLE-2007¹ lies in its focus on identifying relevant intervention targets, contributing to the determination of appropriate treatment intensity levels, assessing offender progress in addressing their treatment needs, and monitoring on-going risk potential risk levels. Feedback from front line users suggests that the introduction of dynamic risk assessment tools into their standard practices has improved their perception of their work effectiveness. Nicholls et al. (2010) noted that police and probation staff valued the use of STABLE-2007 because they felt it improved their assessment skills and enriched their risk management practices. They reported an increased awareness of the most important criminogenic issues which allowed for more targeted responses, particularly to high-risk offenders, during supervision. Professionals using STABLE-2007 have indicated it allows them to focus on the empirically supported risk relevant areas, and track positive client progress in those domains, both of which contribute to improved organization and prioritization of cases on their case load (Ryan, Wilson, Kilgour, & Reynolds, 2014; Walker & O'Rourke, 2013; Watson & Vess, 2007). Despite user endorsements for STABLE-2007's clinical utility and its empirically supported predictive validity, a continuing challenge to those who use the STABLE-2007 is the assessment of individuals in custodial settings and particularly those serving lengthy sentences.

1) Although the present paper is focused on STABLE-2007, given the revisions from STABLE-2000 to STABLE-2007 (Hanson et al., 2007), were relatively limited (slightly increased structure for scoring Capacity for Relationship Stability and Sexual Deviance items and dropping of attitude items) the research on STABLE-2000 is presumed to have relevance to STABLE-2007. As such, some mention of STABLE-2000 is included in the present document where relevant.

Completing STABLE-2007 assessments in an institutional environment poses a number of challenges. These limitations to institutional assessments of dynamic risk factors can foster legal arguments where the defence claims that there is “no evidence of current risk” and the prosecution claims there is “no evidence of change”. Evaluators feel less comfortable making definitive statements about change and clients feel frustrated and demoralized when their perception is they have worked hard to adjust and adapt their behaviours, and this is not reflected in risk assessments.

The present document provides insight into some of the challenges faced by practitioners scoring STABLE-2007 in an institutional environment, particularly noting the added difficulties of assessing someone who has been incarcerated for a long period of time. An overview of the available research on the interrater reliability, construct validity, and clinical utility of STABLE-2007 when scored in institutional settings is provided. Tips for scoring STABLE-2007 in institutional settings include incorporating the concepts of offence analogue and offence replacement behaviours, expanding assessment information sources, and considerations relevant to the weighting of historical versus more recent information into the scoring are outlined. The conclusion is that although more research is needed, particularly for scoring individuals with lengthy sentences, there is currently sufficient research to support that STABLE-2007 can be reliably scored and have clinical utility within an institutional setting. Several recommendations are made to ensure high quality assessments for practitioners in institutional settings.

Challenges in Institutional Settings

Edens and Boccaccini (2017) provide a comprehensive overview of potential differences in scoring of forensic instruments in research versus clinical settings, many of which are relevant to institutional settings. The authors note the possibility of differences in access to file or collateral information, the absence of reliability checks in clinical work, the degree to which interview information is incorporated, differences in informed consent procedures, confidentiality, and the right to refuse participation, differences in the potential outcomes (stakes) for the person being evaluated, as well as the potential for adversarial allegiances among evaluators as key differences that could impact scoring.

Institutional environments are, by definition, more structured, restricted, and confining, which means that the information and data available for scoring items may be significantly limited. This is partly due to fewer opportunities for the individual to engage in or demonstrate behaviours that might be considered reflective of increased or decreased dynamic risk within these environments. Additionally, the functions and qualifications of those who interact most with and contribute to file information about clients adds incrementally to the challenge. Typical sources of collateral information found on institutional files include both professionals and paraprofessionals reflecting a wide range of roles, backgrounds, training, and qualifications. Examples include correc-

tional officers, work supervisors, treatment providers, case managers, and sometimes support staff such as cleaners and food service workers. The day-to-day work-related tasks of such individuals often do not require them to think about dynamic risk factors in the same way that evaluators both prefer and need. Even with specific training on understanding and observing dynamic risk it is difficult to ensure that those involved with clients on a daily basis consistently detect, let alone record things they would not naturally notice or are not directly relevant to the demands of their routine work tasks (Hausam, Lehmann, & Dahle, 2018). Various collateral sources may have different ‘lenses’ through which they see clients. Not only are behaviours likely context specific but what a correctional officer expects from individuals under their charge may be quite different from what someone like a treatment provider might expect (Atkinson & Mann, 2012).

Additionally, some behaviours are more easily and accurately observed or interpreted than others by staff. The salience of bad behaviour over good behaviour is a well accepted concept within the social psychology literature (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Thornton (1985) examined how prison officers perceived inmates and then compared those perceptions with the offender’s self-report data. Prison officers’ ratings of belligerent and uncooperative behaviour and social withdrawal correlated well with the self-reported attitudes of incarcerated men, but officers’ ratings of psychological distress were poorly correlated with the incarcerated men’s self-report, suggesting indications of distress may be less salient to prison officers than oppositional behaviours. Low rate behaviours that are not observed regularly may be another stumbling block when relying upon file information. Research in other types of institutional settings (i.e., residential facilities for the developmentally disabled) have noted that high frequency behaviours are more likely to be noticed, whereas low frequency behaviours are easily missed (Reis, Wine, & Brutzman, 2013). As a result, observers are less likely to observe or record low frequency behaviours let alone the absence of a behaviour. There is good reason to believe this is just as true in secure/correctional institutions (Atkinson & Mann, 2012). Unfortunately, as an evaluator it is difficult to know if the absence of a behaviour in the record means it did not occur or if it means it was simply not observed or observed and not recorded.

The difficulty in obtaining high quality relevant information related to institutional functioning may lead some evaluators to lean more heavily on pre-incarceration file information. While there is some rationale to this approach (i.e., the best predictor of future behaviour is past behaviour) it likely undervalues the impact of changes that occur simply with the passage of time (Hanson, Harris, Helmus, & Thornton, 2014) or as a result of efforts to address criminogenic needs. After a long institutional period, historical behaviours from the community become less relevant because the world, and the individual’s role in the world, change. Someone who is incarcerated at 25 years of age and released at 50 years of age is not going to return to the same social role, family context, pressures, or demands that were in place when they were initially incar-

cerated. Circumstances and behaviour prior to incarceration may have little relevance to post-release contexts. For example, aging itself can result in decreased sexual drive, a particularly important dynamic risk factor for men who have committed sexually motivated offences. In addition, changes in dynamic risk as a function of treatment participation while incarcerated has also been shown to decrease risk as compared to risk assessments based on pre-incarceration history (Sowden & Olver, 2017). Reliance on dated file information may impact the validity of the measure and how reliably the measure is scored.

Interrater Reliability

An important first step to determining STABLE-2007's clinical usefulness is ensuring it can be reliably scored. Given the impact decisions based on assessments that include a STABLE-2007 score may have on the life of the person being assessed, it is important to understand both whether the STABLE-2007 can be scored reliably in general but also the extent to which we can expect consistency across raters under the "real life" clinical conditions of scoring STABLE-2007 within an institution. Coding rules for STABLE-2007 (Fernandez, Harris, Hanson, & Sparks, 2014) state that the measure should be scored based on both an in-depth interview and a comprehensive file review. Interviews are not always feasible when conducting research but in clinical settings interviews are typically a staple of the assessment process. On the positive side, under research conditions chosen raters may have good knowledge of the measure and more generally of psychometric principles. On the negative side, in the absence of an interview they are limited to the quality of available file information, and may have less experience with the files, the available information within the files, and the contexts within which that information is obtained. In clinical settings where the STABLE-2007 is used as part of a comprehensive assessment for decision making it is typically scored by front line clinical staff. While these staff may have little or no training in psychometrics, by conducting their own interviews they may have access to enhanced information not included in files, and typically have a better understanding of the contexts within which records are created and maintained. It is possible that this could contribute to a more thorough and therefore accurate assessment of the individual and may help us understand the circumstances under which accurate assessments are maximized.

The most frequently reported interrater reliability statistic is the Intraclass Correlation Coefficient (ICC). For interpreting ICC values Cicchetti and Sparrow (1981) cite the following guidelines: values of .40 to .59 are considered fair, values of .60 to .74 are good, and values of .75 and above are considered excellent. In a recent meta-analysis evaluating the predictive and incremental validity of STABLE-2007 Brankley et al. (2021) noted that interrater reliability was reported in four studies with the median ICC for STABLE-2007 being .90 (range = .38-.92). The interrater studies noted by Brankley et al.

(2021) included Eher et al. (2012), Hanson et al. (2007), Sowden and Olver (2017), and Webb et al. (2007). These four studies plus two additional studies, Fernandez and Helmus (2017) and Nicholls et al. (2010) are divided into community versus institutional settings and their interrater reliability findings are described in more detail below.

In community settings, most studies reporting interrater reliability of the STABLE-2000 and STABLE-2007 have been scored under research conditions. The original Dynamic Supervision Project (Hanson, Harris, Scott, & Helmus, 2007) which assessed subjects on probation in the community, reported an ICC of .89 for Stable-2000 (the earlier version of STABLE-2007) based on 87 cases. Webb, Craissati, and Keen's (2007) study assessed individuals in the community on probation for convictions of either child pornography or offences against children. The interrater reliability check included 10 randomly selected cases with raters obtaining acceptable levels of agreement (90% correct) on an "updated" version of STABLE-2000 recommended by the authors (i.e., STABLE-2007). Nicholls et al. (2010) evaluated interrater reliability among probation officers in England and Wales using STABLE-2007 to assess clients in the community. Ten files were rated by at least two probation officers and two "expert" raters (i.e., STABLE-2007 trainers). Low to moderate agreement was found among the Stable items and moderate agreement on the total scores. While the results are generally encouraging, in all three of the above studies the authors noted that the scores were based solely on file information and that the quality and depth of information available in each case varied. Hanson et al. (2007) further cautioned that in their study the second raters were not blind to the previous ratings and that the second rater was able to question the original rater about information that was missing or ambiguous, potentially inflating reliability.

Relevant to institutional contexts, Sowden and Olver (2017) double coded 21 randomly selected files from their study of a high-risk incarcerated sample and reported a disappointing but fair interrater reliability (ICC = .46) for STABLE-2007 at pre-treatment after an outlier was removed. The pre-treatment ICC including the outlier was .38, however, the outlier was described as an "early" ICC rating case and the authors believed that removal of the outlier was more representative of the overall interrater reliability. The ICC for post-treatment ratings was .61 (good) and the ICC for change ratings on the STABLE-2007 was .83 (excellent). The authors noted that coders found some STABLE-2007 items (e.g., significant social influences) challenging to rate on an incarcerated sample. Additionally, coding was completed based solely on file information and the authors acknowledged that likely some information required to code key constructs was missing or suboptimal. It is possible that the quality and density of information collected during treatment and therefore available to inform both post-treatment and change ratings improved the interrater reliability for post-treatment ratings and contributed to the solid interrater reliability of change ratings. If accurate this underscores the importance of access to comprehensive information in scoring STABLE-2007.

Studies of interrater reliability of the STABLE-2000 and STABLE-2007 reflective of clinical conditions appear to have fared better. The Federal Evaluation Center for Violent and Sexual Offenders (FECVSO) in Austria reported interrater reliability of STABLE-2000 and STABLE-2007 as part of a series of studies based on their population of adult males released following incarceration for pedosexual offences². Eher et al. (2012) indicated that study subjects were assessed while incarcerated and had a mean sentence length of 32.3 months ($SD = 22.0$) with the mean duration between date of assessment and release being 16.7 months ($SD = 9.2$). Given the length of incarceration prior to the completion of the assessment the authors believed that the item Significant Social Influences could not be accurately scored and it was therefore omitted from scoring. This does, however, indicate that scoring of the other items included information about behaviours assessed during incarceration. Further, the STABLE-2000 and STABLE-2007 were rated by two independent clinicians who conducted their own interview and were blind to the other clinician's ratings, which is a more robust test of the measures' reliability in day-to-day practice. Positively, total score ICCs of .89 for STABLE-2000 and .90 for STABLE-2007 were reported based on 10 randomly selected cases (Etzler, Eher, & Rettenberger, 2020; Rettenberger, Matthes, Schilling, & Eher, 2011) and 15 randomly selected cases (Eher, Matthes, Schilling, Haubner-MacLean, & Rettenberger, 2012).

Fernandez and Helmus (2017) looked at the interrater reliability of STABLE-2007 scoring in an institutional setting under clinical, rather than research conditions. In this study the STABLE-2007 was scored independently by two evaluators using the same file information but conducting independent interviews and blind to each other's scores. Fifty-five consecutive cases of adult males serving sentences of 2 years or more for sexual offences were assessed. The average sentence was 4 years with a range of 2 years to a life sentence, with 15% serving a sentence of 8+ years. Those serving sentences of 8+ years had typically been incarcerated for 2 or more years prior to arrival at the assessment unit. Agreement between the two independent raters was exceptionally high for total scores ($ICC = .92$). All items were in the excellent range except for Significant social influences ($ICC = .56$), Negative Emotionality ($ICC = .70$) and Emotional Identification with Children ($ICC = .70$). Percent agreement was generally good with 8 of 13 items having identical scores 70% + of the time. Qualitative information obtained from the raters indicated that the lower interrater reliability on Significant Social Influences, Negative Emotionality and Emotional Identification with children appeared to be due to discrepant information provided during the independent interviews (Significant Social Influences) and dissimilar presentations by the participants (Negative Emotionality, Emotional Identification with children). However, discrepant total scores resulted in different nominal risk/need categories in only four out of 55 cases. The above studies suggest

2) Given that the data collection period for two of the studies was prior to the introduction of STABLE-2007, the two items that changed from STABLE-2000 to STABLE-2007 were scored retrospectively.

that reliable scoring is possible within an institutional setting, while acknowledging that some items may be more challenging to score in this context. Also, it appears that the inclusion of interview information contributes to improved interrater reliability.

Although the [Fernandez and Helmus \(2017\)](#) study suggests that scoring of some items may be differentially affected by the presentation of the person being interviewed, this finding underscores the importance of considering multiple sources of collateral information, and not just self-report, when scoring items. Further, [Fernandez and Helmus \(2017\)](#) add that their positive results may have at least been partially due to the significant attention paid to ensuring high quality evaluations within this unit. All raters were trained by a certified trainer, were quite experienced in using the measure, had regular access to an expert scorer, and were provided consistent and supportive supervision. This may explain the high interrater reliability but also highlights the benefits of good training and supervision. Other jurisdictions have also noted the benefits of consulting with colleagues and mentors when scoring the STABLE-2007 ([Walker & O'Rourke, 2013](#)).

The available studies that report interrater reliability information for scoring STABLE-2000/2007 in institutions suggest good to excellent levels of reliability. However, information on how long the sample was incarcerated prior to assessment was found for only one study ([Fernandez & Helmus, 2017](#)) and none of the studies reported interrater reliability specific to scoring for individuals who have been incarcerated for long periods. While the existing data is promising in that it indicates STABLE-2007 can be reliably scored in an institutional setting where institutional behavioural information is integrated into the scoring, more research is needed specific to the interrater reliability of scoring for cases that have been incarcerated for lengthy periods (e.g., 10+ years).

Construct Validity and Clinical Utility in Institutional Settings

Evidence for the predictive validity of the STABLE-2007 is provided in the recent [Brankley et al. \(2021\)](#) meta-analysis. Beyond predictive validity, items on the STABLE-2007 are only clinically useful if they do in fact reflect the issues or features they are intended to assess and are consistent with existing theories of dimensions relevant to sexual offending when scored on individuals in an institutional setting. Relevant to this question, [Nunes and Babchishin \(2012\)](#) examined the construct validity of the STABLE-2000 and STABLE-2007. They examined correlations between selected items and validated independent measures of relevant constructs in samples of incarcerated individuals convicted of sexual offences. For the 33 subjects included in the study the average sentence was just under 5 years, although seven subjects had life sentences. Either the STABLE-2000 or STABLE-2007 was scored by psychologists and trained program officers as part of the risk assessment process at one or more points during the sentence. When more than one assessment was available, the most recent was used, suggesting

that institutional information was included in the scoring materials. For Lovers/Intimate Partners, General Social Rejection/Loneliness, Rapist Attitudes, and Child Molester Attitudes items of the STABLE-2000 shared 4% to 19% of the variance with self-report measures of intimacy, loneliness, beliefs supportive of rape, and beliefs supportive of child molestation. The Deviant Sexual Interests/Preferences item from the STABLE-2007 shared significant variance (66%) with the Screening Scale for Pedophilic Interests (SSPI) and 8% with the Sexual Interest Profiling System (SIPS). The results generally suggested that items are associated with measures of similar constructs.

Sowden and Olver (2017) found convergence between the STABLE-2007 and the Violence Risk Scale – Sexual Offense version (VRS-SO), which is a parallel risk measure that includes an assessment of dynamic change. Pre- and post-treatment scores on both measures predicted nonsexual violent, any violent (including sexual), and general recidivism after controlling for static (using Static-99R) risk. Interestingly, STABLE-2007 pre- and post-treatment scores did not predict sexual recidivism in this study, although VRS-SO scores did. While both scales reflected pre-post differences (change scores) of about $\frac{3}{4}$ of a standard deviation, only the VRS-SO change scores were significantly associated with reductions in nonsexual violent, any violent, and general recidivism, but not sexual recidivism. STABLE-2007 change scores did not significantly predict reductions in any recidivism outcomes. Sowden and Olver (2017) suggest the possibility that the STABLE-2007 is less sensitive to change among an incarcerated population given it was developed on a community probation and parole sample while the VRS-SO was developed on an incarcerated treated sample. They note that item operationalization and scale construction is likely informed by characteristics specific to the development sample. While this is a plausible explanation it is worth noting that Hanson et al. (2007) found that the most recent of three or more STABLE-2007 scores was the best predictor of recidivism scores compared to first score, worst score, and rolling average score. Multiple assessments have the advantage of accounting for regression to the mean that may occur across two assessments and consequently may be a better reflection of true change than change scores based on two (a pre- and post-treatment) assessment.

Etzler et al. (2020) examined the underlying dimensions of STABLE-2007 within their population using factor analysis. The analysis yielded three factors deemed to reflect Antisociality, Sexual Deviance, and Hypersexuality. The first factor included seven items reflecting Antisociality, incorporating the items capacity for relationship stability, hostility toward women, general social rejection/loneliness, lack of concern for others, impulsive acts, poor cognitive problem solving, and negative emotionality/hostility. Sexual deviance included two items, namely emotional identification with children, and deviant sexual interests. Hypersexuality also included two items considered to reflect sexual self-regulation difficulties characterized by the items sex drive/preoccupation, and sex as coping. Antisociality and Sexual Deviance were found to be significant predictors of sexual recidivism while Sexual Deviance was negatively associated with non-sexual

violent recidivism. These findings are consistent with previous research demonstrating that static risk variables can be assigned to antisociality or sexual deviance dimensions (Hanson & Morton-Bourgon, 2004, 2005) as well as models of sexual offending that emphasize these two dimensions (Lalumière, Harris, Quinsey, & Rice, 2005; Seto, 2008).

In a study of the VRS-SO and its factor structure, Olver et al. (2018) found strong correlations with the STABLE-2000 scale, individual items, and the three identified VRS-SO factors (Sexual Deviance, Criminality, and Treatment Responsivity). The authors concluded that the two scales assess similar item content and dynamic need areas for both pre- and post-treatment assessments. In particular, strong associations were found between the two scales for Deviant Sexual Interests and Deviant Sexual Preference, Sexual Preoccupation and Sexual Compulsivity, and the Sexual Self-Regulation subscale and the Sexual Deviance Factor.

Seto and Fernandez (2011) examined if men incarcerated for sexual offences assessed on the STABLE-2000 in an institutional setting could be assigned to different dynamic risk groups consistent with theories of adult sexual offending. Based on 419 cases they identified four dynamic risk groups including: a *low needs* group who scored below the overall sample mean on all items; a *typical* group who had intermediate scores across all items; a *sexually deviant* group who scored relatively high on deviant sexual interests, sexual preoccupation, emotional identification with children, and child molester attitudes; and a *pervasive high-needs* group who scored relatively high on many items, reflecting a variety of problems in both general and sexual self-regulation. Interestingly, the dynamic risk groups were not redundant with victim characteristics or offence type. Although, theories of sexual offending often refer to the import of antisociality and sexual deviance (Brouillette-Alarie et al., 2016), there was no clear antisocial group identified. The closest was a pervasive high needs group who scored relatively high on both antisocial and sexual deviance items. However, the lack of redundancy between dynamic risk groups and offender types did reflect previous research demonstrating higher than expected sexual deviance among groups typically expected to be more antisocial than sexually deviant such as convicted rapists (Lalumière et al., 2005) and groups generally expected to be less deviant such as incest offenders (Fernandez, 2001). The authors suggest that developing typologies of sexual offending persons based on dynamic risk factors may provide a better understanding of the psychological origins of sexual offending over typologies based on victim choice. Further, from a clinical utility perspective, the authors suggest that designing interventions based on dynamic risk assessment may be more effective than focusing on victim choice (e.g., rapists versus child molesters).

While the research is both limited and preliminary, to date the available studies generally support the use of the STABLE-2007 in institutional settings as a reliable and clinically useful measure. The current evidence suggests that STABLE-2007 scored in an institutional setting can be scored reliably under both research and clinical conditions,

that the scale and items are associated with other measures of similar constructs validated within institutional settings, that resulting profiles based on institutional scoring reasonably reflect existing theories of sexual offending, and that these profiles may be more relevant to intervention design than other typologies.

Tips for Using STABLE-2007 in an Institutional Setting

There are several significant challenges to institutional scoring of STABLE-2007. One challenge is related to evaluating behaviours within a heavily structured environment where opportunities for demonstration, observation, and recording of behaviours may be more limited. A related challenge is assessing change in this type of structured environment, where both problematic behaviours, but also behavioural improvements, may manifest differently than they do in the community. Additionally, both of these issues are impacted by how the evaluator chooses to weigh historical versus more recent behavioural information in scoring.

In completing a comprehensive assessment within an institutional setting, the difficult work often includes finding and evaluating information. Ideally, item scoring should be based on a confluence of assessment strategies including a comprehensive clinical interview, relevant psychological testing, review of available documentation, and collateral reports. In an institutional setting item scoring necessarily requires interpreting whether the available information suggests the presence of risk factors or reflects alternative, prosocial behaviours, or protective factors. Researchers have had some success looking at common institutional behaviours as predictors of future violence. For example, institutional infractions have been shown to predict post-release aggression (Cochran, Mears, Bales, & Stewart, 2014; Mooney & Daffern, 2011). However, researchers note that the data is “muted” compared to behaviours observed among individuals in unconstrained environments (Mooney & Daffern, 2011), which seems likely also to be the case for dynamic risk factors relevant to sexual offending and makes interpretation of those behaviours challenging. Consequently, as mentioned previously, critical to interpreting institutional data is understanding the files, the available information within the files, and the contexts within which that information is obtained.

Offence Paralleling or Offence Analogue Behaviours (OABs) and Offence Replacement Behaviours (ORBs)

Extremely helpful to this task are recommendations that clinicians think of institutional information regarding client behaviours through a lens that reflects the different realities

of institutional life. These include the consideration of Offence Paralleling Behaviours (OPBs) (Daffern, Jones, & Shine, 2010) or Offence Analogue Behaviours (OABs) and Offence Replacement Behaviours (ORBs) (Gordon & Wong, 2015). Offence Paralleling Behaviours (OPBs) are defined as a “behavioural sequence incorporating overt behaviours (that may be muted by environmental factors) appraisals, expectations, beliefs, affects, goals and behavioural scripts, all of which may be influenced by the patient’s mental disorder, that is functionally similar to behavioural sequences involved in previous criminal acts.” (Daffern et al., 2007, p. 267). In essence, the evaluator is directed to look for behaviours that, when considered within the realities of an institutional environment, still appear to represent previously seen behaviours relevant to the individual’s criminal behaviour while in the community. Engaging in fights with other inmates, quitting rehabilitation programs, or running up institutional debts reflect institutional specific examples of dynamic risk factors such as impulsivity and poor problem solving. Gordon and Wong (2015) go further by not only recommending evaluators consider what they refer to as OABs, or behaviours within a custodial setting that indicate manifestations of dynamic risk factors (similar to Offence Paralleling Behaviors) but also to ORBs, described as risk-replacement or pro-social behaviours that are observable in custodial or other types of controlled environments. An example might include a situation where the individual faces a significant disappointment or stressor, such as being denied conditional release, and responds with appropriate coping strategies such as seeking support and using relaxation or self-soothing skills such as meditation or listening to music. It is recommended that the presence of OABs suggest the problem behaviour (considered a proxy for the risk factor) still exists, while an absence of OABs indicates the possibility of change. The presence of ORBs in contrast is considered clearer evidence of change. The longer the ORBs have been present, particularly within high risk contexts, the more confident the assessor can be of substantive change to risk relevant propensities. In both cases, evaluators are encouraged to consider how evidence of the presence of behaviors relevant to the risk factors, or evidence of alternative behaviours contrary to those risk factors, can be seen within the limited, and in some ways artificial, institutional environment.

Olver, Gordon, and Wong (2017) have developed an OAB/ORB guide for sexual offending populations adapted to the VRS-SO, but the guidance may be applied to other tools, including the STABLE-2007. Olver and Stockdale (2020) provide a more general overview of the concepts of OABs and ORBs and discuss how these types of behaviours can be monitored and changed within an institutional environment. Certainly, approaching institutional information and behavioural observations with a behavioural “paralleling” mindset may help expand the evidence both for and against dynamic risk factors during an institutional assessment. Even in prison or hospital settings, there are opportunities for a range of both antisocial and prosocial behaviours that can be useful to scoring dynamic risk items. Further, while staff do not always attend to risk relevant

information in the way a risk evaluator might hope, files often include information that can be used to inform OABs and ORBs, even if those recording the information did not have such concepts in mind. This does, however, require significant work by the evaluator and may be more difficult depending on the setting or other demands such as time pressures. Regardless, evaluators using STABLE-2007 within institutional settings are encouraged to read about the above concepts and consider their application to scoring STABLE-2007.

Expanding Information Sources

Realistically, some STABLE-2007 items present as easier to score within an institutional setting than others. Capacity for relationship stability, hostility toward women, impulsive acts, poor problem solving, negative emotionality, and cooperation with supervision are all items where either information may be relatively easy to find and verify (e.g., capacity for relationship stability) or opportunities to demonstrate behaviours that suggest either continued risk or new alternative behaviours are both reasonably frequent and have more potential for being captured in interactions and reports by other staff. Evidence of change such as reports of improved behaviours or even decreased reports or an absence of reports of problematic behaviours over time, are potentially quite meaningful in an institutional setting. In contrast items that are either dependent on interactions that are not generally available within institutions (e.g., significant social influences, emotional identification with children, general social rejection, lack of concern for others) or reflect behaviours that most commonly occur in private (sex drive/preoccupation, sex as coping, deviant sexual preferences) may be more difficult to assess. Comprehensive assessments of these items may require going beyond the standard file review and interview to collecting collateral information from sources outside of the direct institutional environment. For example, while institutional social circles may be relevant to scoring significant social influences, it may also be helpful to ask for permission to speak to those the client identified as influential in their life outside of the institution, especially if these individuals are part of their release plans. This provides the opportunity to both directly assess that person's involvement and contact with the client as well as evaluate whether they are a positive or negative influence. Additional options include confirming who is identified as an emergency contact in institutional files, checking institutional phone and correspondence logs to determine frequency of contact, or even asking the client to provide examples of their correspondence with their identified significant social influence. These types of collateral investigations may also be relevant to scoring the general social rejection and lack of concern for others items, where a clearer understanding of the client's interactions with those in his social networks is important to scoring the item.

Likely the most challenging items to score in an institutional setting are the sexual self-regulation items (sex drive/preoccupation, sex as coping, sexual deviance). Although some problematic behaviours may present institutionally for those with severe sexual self-regulation deficits, in most cases these items rely heavily on self-report. Further the information is not easy to verify because it is based on internal processes (e.g., thoughts and fantasies) or behaviours that typically occur in private. This is, however, true for both institutional and community assessments and speaks to the critical importance of building rapport, ensuring a collaborative working relationship with the client, and clearly explaining the benefits to the client of accurately identifying risk areas for intervention. Further there are possibly additional sources of information on these items that may be available institutionally that are unlikely to be found among community samples. Given the relative lack of privacy within institutions staff and roommate complaints about inappropriate or high frequency sexual behaviours are more likely to be reported than in community settings. This includes sexualized coping in response to negative events, especially if staff are trained to notice and record this type of behaviour. While unfettered access to child victims does not occur within institutions there may be reports available from contexts where there are opportunities to observe behaviours, such as staring or attempting to interact with the children in institutional visiting areas. Finally, in jurisdictions with conjugal visits, permission to speak with the client's sexual partner may provide additional insights.

To some degree limitations to institutional scoring of STABLE-2007 items may be at least partly addressed by the thoroughness and creativity of the evaluator's investigatory processes. Of note, this is likely impacted by the evaluator's understanding of the nature of the institutional environment and the resultant availability of opportunities for both risky and alternative behaviours. As an example, it is important to understand the reality of access to intoxicants within an institution to evaluate the client's coping skills in this area. If the evaluator assumes intoxicants are not available within an institution, then evidence of control strategies being used by the client will be missed.

Weighting Historical Versus More Recent Information Sources

An advantage to scoring STABLE-2007 for men who have been incarcerated for long periods may include access to voluminous information. The challenge then becomes how to balance historical information, including pre-incarceration information, with the more recent information as discussed above, which may evidence change. For those with long sentences it is not uncommon to have file information detailing poor pre-incarceration (community) behaviours that continue into the early years of the client's sentence. After a significant period of incarceration, and potentially participation in treatment interventions, there may be little evidence of those original problem behaviours and even

evidence of alternative, prosocial, behaviours within the institution but it is unclear how to balance this continuum of information.

The current STABLE-2007 Coding Manual (Fernandez et al., 2014) provides little in the way of guidance regarding how to weight historical versus recent information. Page 23 of the coding manual notes “Items scores are based on both interview and collateral information. The ratings represent estimates of the individual’s typical or current ‘baseline’ functioning. Assessors should consider recent and historical behaviours; however, the primary task is determining expected functioning over the next six to 12 months.” While it is suggested that evaluators “consider” both recent and historical behaviours, there are no specifics as to how to weight these two, possibly contradictory, sources of information. Of relevance to this issue is Olver et al.’s (2018) finding that for the VRS-SO both pre-treatment dynamic risk and change scores made independent contributions to predicting risk. This might suggest that a defensible approach to any dynamic risk assessment is equally weighting historical and recent behaviour. The negative to this approach, however, is an individual’s risk will never be reduced past a certain point if historical information, some of which may be decades old at the time of assessment, is given equal weighting across all assessments indefinitely. That is, an item that is initially scored a “2” can never be reduced to a “0” if historical information is always given equal weighting. Consequently, there are some considerations in principle that could be used to modify this weighting. These additional considerations include: 1) the persistence and recency of historical behaviour patterns. Chronic problems that continue to be evident in the more recent past (e.g., past year or two) may indicate continued vulnerabilities in this area; 2) Evidence of both intellectual and behavioural changes. The client should be able to acknowledge and demonstrate an understanding of the problematic behaviours and be able to articulate what changes they have made in addition to demonstrating clear behavioural changes; 3) The duration of the more recent behaviour changes. The longer period a behavioural change has been evident, the more likely the change has been integrated into the client’s skill set; and 4) The degree to which the current setting resembles the community. Evidence that behaviour has changed within contexts where behavioural opportunities are more unfettered are stronger indications of a well integrated and enduring skill set. While reductions for items of scores from a “2” to a “1” may be based on the first three considerations (e.g., no recent evidence of problematic behaviours and evidence of intellectual and behavioural changes sustained over a reasonable period of time) a score reduction from a “2” to a “0” would likely require evidence that the behavioural change remains even when the individual is faced with unconstrained opportunities to return to their original behaviour patterns. The STABLE-2007 development team is currently working on providing additional guidance regarding the definable conditions under which a score of “2” can change to a “1” or “0”.

Ensuring High Quality Assessments

Reliably scoring a measure of risk has also been linked to the type and possibly quality of training provided, with training by experienced credible peer trainers associated with improved reliability over training by an expert developer (Vincent, Guy, Fusco, & Gershenson, 2012). The authors suggest that experienced peer trainers may be better positioned to address scoring issues that are idiosyncratic to the setting or population during the training. It is also possible that learning from experienced peer trainers inspires confidence in participants that expertise in test development is not necessarily critical to good scoring.

Two strategies have been associated with substantial increases in predictive accuracy of measures. The first is involvement of the scale's developer (Andrews et al., 2011). The scale developer, who is typically familiar with the scoring of the scale and invested in its implementation, is assumed to boost fidelity to the scale during implementation. This also may be due to bias or allegiance effects to the scale. For obvious reasons this is not an issue in most clinical or institutional settings, as front-line clinicians rarely have a strong association with a particular measure. The second strategy suggests expanding this scoring "investment" or "commitment" often found among scale developers to front-line users. In the original Dynamic Supervision Project the potential effect of commitment was evident when it was found that the predictive accuracy of STABLE-2000s scored by a subgroup of "conscientious" probation officers was higher than the combined group of all probation officers. The defining feature of the "conscientious" group was that they completed all the steps that were requested of them in data collection. This consistent adherence to the identified requirements of the study suggested a personal commitment to the process among this group that ultimately resulted in better, more accurate assessments (Hanson et al., 2007).

Recommendations and Conclusion

To date the available studies generally support the use of the STABLE-2007 in institutional settings as a reliable and clinically useful measure. However, the variability in study results does indicate that attention to the conditions under which scoring is completed, how file information is interpreted, interviewing techniques, and diligent gathering of collateral information may impact the reliability and utility of the measure. This is likely true of any assessment tool that assesses dynamic risk factors. As a result, the following strategies for ensuring high quality STABLE-2007 scoring are recommended.

1. Ensure comprehensive information gathering and scoring considerations. As noted previously the commitment and conscientiousness of the evaluator can have a significant impact on risk assessment accuracy.
 - Consider how risk relevant behaviours, both positive (i.e., evidence of improved skills) and negative (evidence of continuing or worsening problems) may be manifested in institutional settings. Reviewing the concepts of OABs and ORBs (Olver, Gordon, & Wong, 2017) may be helpful.
 - Expand information gathering beyond the most commonly available file and interview information. Collect reports from additional sources where possible, such as work, school, program, and visitation contexts. Make direct contact when possible, with significant collateral contacts.
 - Invest time in developing a solid rapport and a collaborative working relationship with the individual being assessed. The quality of information they provide will impact the accuracy of the assessment. Time spent ensuring a good working relationship will likely pay dividends in the type and amount of information provided at interview.
 - Balance the weighting of historical versus recent information to capture change over time. In particular give consideration to 1) the persistence and recency of historical behaviour patterns; 2) evidence of both intellectual and behavioural changes; 3) the duration of more recent behaviour changes; and 4) the degree to which the current setting resembles the community.
2. Ensure evaluators have solid knowledge of the instruments through quality training, and sufficient environmental support to promote consistent and accurate scoring.
 - Seek high quality training from a certified trainer. The individual providing training should have experience, know the measure well, and be able to provide guidance on situations that may be unique to the organization or target population.
 - Secure top down support and concentrated efforts to encourage “buy in” and understanding from all staff involved with offenders in order to enhance the quality of the information available in records. Training on risk measures for those who interact with clients on a daily basis within the institution, even if they do not score the measure, will provide insight into the types of information that evaluators need for scoring and hopefully improve recording of risk relevant behaviours.
 - Develop mentorships with those who are more experienced in using the measure available so that novice scorers have an identified person with whom they can discuss their cases and their risk scoring. Many basic questions are easily answered by an experienced scorer. Easy access to an experienced scorer can address mistakes early on before they become a habit.

- Access clinical supervision by a very experienced assessor for tricky questions. Given clinical sites are often under considerable time pressures for completing assessments, evaluators should have access to someone with sufficient knowledge of the measure, site, and population to address complex questions. This person may well organize regular peer review sessions where complex cases can be discussed and resolved.
- Develop a quality control process, either through regular professional development days, internal supervision by senior employees who are committed to the risk assessment process, or possible “scoring clinics” run cooperatively within organizations.
- Require evaluators to “sign off” on their scoring decisions by signing against a statement indicating they believe they had sufficient information and that they feel confident in their decisions. This can have the effect of increasing the “commitment” the evaluator feels toward the scoring process and consequently result in improved accuracy (Hanson et al., 2007).

To date there is preliminary research to support that STABLE-2007 can be reliably scored and has clinical validity and utility in institutional settings. As of yet there is no definitive research specific to the reliability and validity of STABLE-2007 for institutionalized persons serving long sentences. The reality, however, is that evaluators are expected to assess and make recommendations regarding who poses a threat to public safety, what treatment needs are most likely to result in change, and whether someone has benefited from interventions targeting those needs regardless of sentence length. Consequently, while we recommend STABLE-2007 for institutional scoring, including those serving long sentences, we encourage evaluators completing these assessments to follow the recommendations outlined above and to consider and document cautions about the difficulties in assessing change in controlled environments. Although more research is needed, in the meantime, clinicians can do their part to contribute to the credibility of risk and need assessments by ensuring assessments within institutional environments are comprehensive and high-quality. Ultimately, thoughtful implementation and solid support are believed to result in quality assessments, including within an institution and for those with long sentences.

Funding: The author has no funding to report.

Acknowledgments: The author has no additional (i.e., non-financial) support to report.

Competing Interests: The author has declared that no competing interests exist.

References

- Andrews, D. A., Bonta, J., Wormith, J. S., Guzzo, L., Brews, A., Rettinger, J., & Rowe, R. (2011). Sources of variability in estimates of predictive validity: A specification with level of service general risk and need. *Criminal Justice and Behavior, 38*(5), 413-432.
<https://doi.org/10.1177/0093854811401990>
- Atkinson, D. F., & Mann, R. E. (2012). Prison officers as observers of offence paralleling behaviours: An untapped resource? *Journal of Forensic Psychiatry & Psychology, 23*(2), 139-155.
<https://doi.org/10.1080/14789949.2012.668209>
- Brankley, A. E., Babchishin, K. M., & Hanson, R. K. (2021). STABLE-2007 demonstrates predictive and incremental validity in assessing risk-relevant propensities for sexual offending: A meta-analysis. *Sexual Abuse, 33*(1), 34-62. <https://doi.org/10.1177/1079063219871572>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Brouillette-Alarie, S., Babchishin, K. M., Hanson, R. K., & Helmus, L. M. (2016). Latent constructs of the Static-99R and Static-2002R: A three-factor solution. *Assessment, 23*(1), 96-111.
<https://doi.org/10.1177/1073191114568114>
- Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*(2), 127-137.
- Cochran, J. C., Mears, D. P., Bales, W. D., & Stewart, E. A. (2014). Does inmate behavior affect post-release offending? Investigating the misconduct-recidivism relationship among youth and adults. *Justice Quarterly, 31*(6), 1044-1073. <https://doi.org/10.1080/07418825.2012.736526>
- Daffern, M., Jones, L., Howells, K., Shine, J., Mikton, C., & Tunbridge, V. (2007). Refining the definition of offence paralleling behavior [Editorial]. *Criminal Behaviour and Mental Health, 17*(5), 265-273. <https://doi.org/10.1002/cbm.671>
- Daffern, M., Jones, L., & Shine, J. (Eds.). (2010). *Wiley Series in Forensic Clinical Psychology: Offence paralleling behavior. A case formulation approach to offender assessment and intervention*. Wiley-Blackwell. <https://doi.org/10.1002/9780470970270>
- Edens, J. F., & Boccaccini, M. T. (2017). Taking forensic mental health assessment “out of the lab” and into “the real world”: Introduction to the special issue on the field utility of forensic assessment instruments and procedures. *Psychological Assessment, 29*(6), 599-610.
<https://doi.org/10.1037/pas0000475>
- Eher, R., Matthes, A., Schilling, F., Haubner-MacLean, T., & Rettenberger, M. (2012). Dynamic risk assessment in sexual offenders using STABLE-2000 and the STABLE-2007: An investigation of predictive and incremental validity. *Sexual Abuse, 24*(1), 5-28.
<https://doi.org/10.1177/1079063211403164>
- Etzler, S., Eher, R., & Rettenberger, M. (2020). Dynamic risk assessment of sexual offenders: Validity and dimensional structure of the STABLE-2007. *Assessment, 27*(4), 822-839.
<https://doi.org/10.1177/1073191118754705>

- Fernandez, Y. (2001). *Phallometric testing with sexual offenders against female victims: An examination of reliability and validity issues* (Unpublished doctoral dissertation). Queen's University, Kingston, Ontario, Canada.
- Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual: Revised 2014*. Unpublished manual, Public Safety Canada, Ottawa, Ontario.
- Fernandez, Y., & Helmus, L. M. (2017). A field examination of the inter-rater reliability of the Static-99 and STABLE-2007 scored by correctional program officers. *Sexual Offender Treatment, 12*(2). Retrieved from <http://www.sexual-offender-treatment.org/181.html>
- Gordon, A., & Wong, S. (2015). The use of a structured guide to assess proxies of offending behaviours and change in custodial settings. *Journal of Forensic Practice, 17*(2), 83-100. <https://doi.org/10.1108/JFP-10-2013-0048>
- Hanson, R. K., Harris, A. J. R., Helmus, L., & Thornton, D. (2014). High-risk sex offenders may not be high risk forever. *Journal of Interpersonal Violence, 29*(15), 2792-2813. <https://doi.org/10.1177/0886260514526062>
- Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (User Report No. 2007-05). Ottawa, Ontario, Canada: Public Safety Canada. Retrieved from <http://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/index-eng.aspx>
- Hanson, R. K., & Morton-Bourgon, K. (2004). *Predictors of sexual recidivism: An updated meta-analysis* (Report No. PS3-1/2004-2E-PDF). Ottawa, Ontario, Canada: Public Safety Canada. Retrieved from <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/2004-02-prdctrs-sxl-rcdvsm-pdtd/index-en.aspx>
- Hanson, R. K., & Morton-Bourgon, K. (2005). The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies. *Journal of Consulting and Clinical Psychology, 73*(6), 1154-1163. <https://doi.org/10.1037/0022-006X.73.6.1154>
- Hausam, J., Lehmann, R. J. B., & Dahle, K.-P. (2018). Predicting offenders' institutional misconduct and recidivism: The utility of behavioral ratings by prison officers. *Frontiers in Psychiatry, 9*(December), Article 679. <https://doi.org/10.3389/fpsy.2018.00679>
- Lalumière, M. L., Harris, G. T., Quinsey, V. L., & Rice, M. E. (2005). *The causes of rape: Understanding individual differences in male propensity for sexual aggression*. Washington, DC, USA: American Psychological Association. <https://doi.org/10.1037/10961-000>
- Mooney, J. L., & Daffern, M. (2011). Institutional aggression as a predictor of violent recidivism: Implications for parole decision making. *International Journal of Forensic Mental Health, 10*(1), 52-63. <https://doi.org/10.1080/14999013.2010.550984>
- Nicholls, C. M., Callanan, M., Legard, R., Tomaszewski, W., Purdon, S., & Webster, S. (2010). Examining implementation of the Stable and Acute dynamic risk assessment tool pilot in England and Wales. In *Ministry of Justice Research Series 4/10*. Retrieved from <https://www.justice.gov.uk/downloads/publications/research-and-analysis/moj-research/sexual-offenders-risk.pdf>

- Nunes, K. L., & Babchishin, K. M. (2012). Construct Validity of Stable-2000 and Stable-2007 Scores. *Sexual Abuse, 24*(1), 29-45. <https://doi.org/10.1177/1079063211404921>
- Olver, M. E., Gordon, A., & Wong, S. C. P. (2017). *Offense analogue and offence replacement behavior (OAB/ORB) rating guide for Violence Risk Scale: Sexual Offense version*. Regional Psychiatric Centre and University of Saskatchewan, Saskatoon, Saskatchewan, Canada.
- Olver, M. E., Neumann, C. S., Kingston, D. A., Nicholaichuk, T. P., & Wong, S. C. P. (2018). Construct validity of the Violence Risk Scale-Sexual Offender version instrument in a multisite sample of treated sexual offenders. *Assessment, 25*(1), 40-55. <https://doi.org/10.1177/1073191116643819>
- Olver, M. E., & Stockdale, K. C. (2020). Assessing and managing offense analogue and offense replacement behaviors in correctional settings. *Advancing Corrections, 10*(3), 42-55. Available from <https://icpa.org/library/advancing-corrections-journal-edition-10-article-3/>
- Reis, M. H., Wine, B., & Brutzman, B. (2013). Enhancing the accuracy of low-frequency behavior data collection by direct-care staff. *Behavioral Interventions, 28*(4), 344-352. <https://doi.org/10.1002/bin.1371>
- Rettenberger, M., Matthes, A., Schilling, F., & Eher, R. (2011). Die Validität dynamisch-veränderbarer Risikofaktoren bei der Vorhersage einschlägiger Rückfälle pädosexueller Straftäter: Eine Studie über Stable-2000 und Stable-2007. *Forensische Psychiatrie, Psychologie, Kriminologie, 5*(1), 45-53. <https://doi.org/10.1007/s11757-010-0086-z>
- Ryan, J., Wilson, N., Kilgour, G., & Reynolds, N. (2014). Balancing efficiency and accuracy: Guidelines for the psychological assessment of extended supervision orders for child sex offenders in New Zealand. *Psychiatry, Psychology and Law, 21*(1), 139-157. <https://doi.org/10.1080/13218719.2013.793154>
- Seto, M. C. (2008). *Pedophilia and sexual offending against children: Theory, assessment, and intervention*. Washington, DC, USA: American Psychological Association. <https://doi.org/10.1037/11639-000>
- Seto, M. C., & Fernandez, Y. M. (2011). Dynamic risk groups among adult male sexual offenders. *Sexual Abuse, 23*(4), 494-507. <https://doi.org/10.1177/1079063211403162>
- Sowden, J. N., & Olver, M. E. (2017). Use of the violence risk scale-sexual offender version and the STABLE 2007 to assess dynamic sexual violence risk in a sample of treated sexual offenders. *Psychological Assessment, 29*(3), 293-303. <https://doi.org/10.1037/pas0000345>
- Thornton, D. (1985). Intra-regime variation in inmate perception of prison staff. *British Journal of Criminology, 25*(2), 138-147. <https://doi.org/10.1093/oxfordjournals.bjc.a047508>
- Vincent, G. M., Guy, L. S., Fusco, S. L., & Gershenson, B. G. (2012). Field reliability of the SAVRY with juvenile probation officers: Implications for training. *Law and Human Behavior, 36*(3), 225-236. <https://doi.org/10.1037/h0093974>
- Walker, M., & O'Rourke, M. (2013). Probation officers' experience of using Risk Matrix 2000 and Stable & Acute 2007 when supervising sex offenders living in the community. *Irish Probation Journal, 10*(October), 162-176. Retrieved from

[http://www.probation.ie/EN/PB/0/9F3B01DD5632A3FA8025802F003D1D86/\\$File/IPJ2013pages162to176Abstract.pdf](http://www.probation.ie/EN/PB/0/9F3B01DD5632A3FA8025802F003D1D86/$File/IPJ2013pages162to176Abstract.pdf)

Watson, T., & Vess, J. (2007). Risk assessment of child-victim sex offenders for extended supervision in New Zealand. *Journal of Forensic Psychiatry & Psychology*, 18(2), 235-247.

<https://doi.org/10.1080/14789940601116267>

Webb, L., Craissati, J., & Keen, S. (2007). Characteristics of internet child pornography offenders: A comparison with child molesters. *Sexual Abuse*, 19(4), 449-465.

<https://doi.org/10.1177/107906320701900408>



Sexual Offending: Theory, Research, and Prevention (SOTRAP) is the official journal of the International Association for the Treatment of Sexual Offenders (IATSO).



leibniz-psychology.org

PsychOpen GOLD is a publishing service by Leibniz Institute for Psychology (ZPID), Germany.