

Assessing the Dimensional Structure and Measurement Invariance of STABLE-2007 Scores in a California Sample of Adult Men on Parole for Sex Crimes

Ryan E. Ditchfield¹ , Allen Azizian¹ , Deirdre D'Orazio² , Lea Chankin³

[1] Department of Criminology, California State University, Fresno, CA, USA. [2] Department of State Hospitals, SVP-Conditional Release Program, Sacramento, CA, USA. [3] California Department of Corrections and Rehabilitation – Division on Adult Parole Operations, Sacramento, CA, USA.

Sexual Offending: Theory, Research, and Prevention, 2026, Vol. 21, Article e19019,
<https://doi.org/10.5964/sotrap.19019>

Received: 2025-07-24 • Accepted: 2026-07-01 • Published (VoR): 2026-09-08

Handling Editors: Martin Rettenberger, Centre for Criminology (Kriminologische Zentralstelle – KrimZ), Wiesbaden, Germany; Sonja Etzler, Medical Center - University of Freiburg, Freiburg, Germany

Corresponding Author: Ryan E. Ditchfield, Department of Criminology, California State University, Fresno, 2576 E. San Ramon, MS/ST 104, Fresno, CA, 93740, United States. E-mail: ditchfield@csufresno.edu

Supplementary Materials: Materials [see [Index of Supplementary Materials](#)]



Abstract

The STABLE-2007 is a 13-item instrument used to identify, treat, and monitor changes in criminogenic needs associated with sexual reoffending, but no empirical research to date has examined the construct validity of this instrument within California. The present study assessed the factor structure and measurement invariance of STABLE-2007 scores in a Californian sample of 873 parolees enrolled in a community sex offense treatment program. Archival scores for the STABLE-2007 were compared across four time points over the course of supervision. Consistent with the broader literature on sexual offending risk, we found preliminary support for an invariant two-factor structure consisting of *antisociality* and *atypical sexuality*. In addition to providing support for the STABLE-2007's construct validity and reliability, our identification of two clear constructs reflecting *antisociality* and *atypical sexuality* lays a foundation for establishing the instrument's validity and reliability within the context of United States community correctional populations.

Keywords

STABLE-2007, sexual offending, risk assessment, construct validity, measurement invariance, exploratory structural equation modeling



This is an open access article distributed under the terms of the [Creative Commons Attribution 4.0 International License](#), CC BY 4.0, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Highlights

- This study examined the construct validity of the STABLE-2007 in a California sample.
- Two sexual offending risk constructs were identified: *antisociality* and *atypical sexuality*.
- These two constructs remained consistent over the course of community supervision.
- These findings extend the validity of this instrument to contexts beyond Canada and Austria.

Accurate interpretation of scores from psychological instruments, including risk assessment instruments like the STABLE-2007, hinges on the underlying validity and reliability of the assessment tool (Flake et al., 2017). Factor analysis is a class of statistical analyses used to establish validity in psychometric instruments (Brown & Moore, 2012; Strauss & Smith, 2009). Factor analysis is an indispensable analytic tool for assessing construct, convergent, and discriminant validity while adjusting for measurement error and method effects. To the authors' knowledge, however, no research has yet assessed the factor structure of the STABLE-2007 within the United States. This means that, although the STABLE-2007 is a commonly required tool for assessment of sexual recidivism risk around the world, the instrument's construct validity has yet to be formally verified in populations beyond Canada and Austria.

To address this gap, the present study assessed the factor structure and measurement invariance for the STABLE-2007 risk assessment instrument within a California sample of male parolees convicted of sexual offenses. The current study draws on modern structural equation modeling approaches to 1) determine the factor structure of the STABLE-2007 within a robust field sample, 2) evaluate the convergent and discriminant validity of identified factors by comparing dynamic and static risk scores, and 3) assess the reliability of the STABLE-2007's factor structure over time. In doing so, this study clarifies the latent dynamic risk factors underlying sexual offending within United States correctional populations, thereby facilitating accurate and effective community supervision and risk assessment of individuals convicted of sexual offenses.

Latent Risk Factors Related to Sexual Offending

Broadly, risk assessment instruments can be categorized as assessing either static factors or dynamic factors that can be changed via intervention (i.e., criminogenic needs). Criminogenic needs, specifically, are enduring yet changeable offender characteristics that represent long-term vulnerabilities related to reoffending and that can theoretically improve in response to treatment (Eisenberg et al., 2019; Mann et al., 2010; Seto et al., 2023). Whereas instruments like the Static-99R evaluate relatively unchangeable static factors (e.g., age, criminal history) as a baseline for recidivism risk, the STABLE-2007 assesses criminogenic needs related to sexual recidivism (Brankley et al., 2021; Phenix

et al., 2017). The STABLE-2007's focus on criminogenic needs is, therefore, designed to address the *Need Principle* of the Risk-Need-Responsivity framework (Bonta & Andrews, 2023; Seto et al., 2023).

Prior theory and research suggest that static and dynamic risk factors for sexual offending can be characterized by at least two constructs representing sexual criminality and antisociality (Brouillette-Alarie et al., 2016). The underlying nature of these factors, however, is dependent on risk assessment instrument characteristics and the specific context of an assessed correctional population. Reliability and validity are properties of scores within a given sample, not general properties of an instrument that can be automatically generalized to new contexts (Cook & Beckman, 2006). This means that revalidation of risk assessment instruments is essential for an instrument's use in new correctional populations, highlighting the strong need for re-assessment of the STABLE-2007's psychometric properties. Identifying the underlying dimensions of risk assessment instruments also has utility with respect to treatment and assessment of recidivism risk, as prior research has shown that general and sexual offending risk factors differentially predict parolee's risk for reoffending (Babchishin et al., 2016).

With regard to instruments assessing static risk, researchers have identified two to three static risk constructs. Brouillette-Alarie and Hanson (2015), for example, found a two-factor solution representing a rule-violation domain and a sex crime-specific domain, but a subsequent examination revealed a three-factor solution reflecting persistence/paraphilia, youthful stranger aggression, and general criminality (Brouillette-Alarie et al., 2016). By contrast, researchers have identified factor solutions ranging from two to four dimensions for dynamic risk assessment instruments. The Violence Risk Scale–Sexual Offending Version (VRS-SO), for example, has consistently produced scores reflecting a three-factor model consisting of sexual deviance, criminality, and treatment responsivity (Olver et al., 2018; Wielinga & Olver, 2024). The dimensionality of dynamic risk assessment instruments has not been fully resolved, however. Different solutions for the VRS-SO exist, including a four-factor solution in two samples of child sexual abusers (Beggs & Grace, 2011). More recent investigations of dynamic risk assessments' construct validity have instead found support for a two-factor model consisting of sexual risk and antisocial opposition (e.g., Sex Offender Treatment Intervention and Progress Scale (SOTIPS); Miner et al., 2023).

Prior Evidence Regarding the Dimensionality and Construct Validity of the STABLE-2007

The initial STABLE-2007 validation study by Hanson et al. (2007) did not specify a factor structure. Rather, STABLE-2007 items were initially selected by 1) empirically identifying supervision issues that predict sexual recidivism among individuals of similar static risk levels and 2) removing items that did not significantly predict sexual recidivism, such as sexual entitlement and attitudes tolerant of rape (Hanson & Harris, 1998, 2000). The

STABLE-2007 coding manual, used as official training material in California, describes its 13 items as falling into five categories: Significant Social Relationships, Intimacy Deficits, Sexual Self-Regulation, General Self-Regulation, and Cooperation with Supervision (Fernandez et al., 2014, p. 14). The validity of this five-dimension model remains untested via peer-reviewed research, however.

Empirical research on the STABLE-2007's construct validity, although sparse, has identified both two-factor and three-factor solutions for the STABLE-2007 (Brien-Robidoux, 2022; Brouillette-Alarie & Hanson, 2015; Etzler et al., 2020). Brouillette-Alarie and Hanson (2015) and Brien-Robidoux (2022) observed two constructs representing anti-sociality and sexual deviance, though the latter study also identified a third uninterpretable factor consisting of only Item 7 (Impulsive Acts). By contrast, Etzler and colleagues (2020) observed a three-factor solution representing antisociality, sexual deviance, and hypersexuality – with the caveat that the authors were forced to exclude Item 1 (Significant Social Influences) scores due to a lack of reliable information about offenders' social status within the Austrian Prison System. Moreover, both Brien-Robidoux (2022) and Etzler et al. (2020) examined the STABLE-2007's construct validity via exploratory factor analysis (EFA) but did not verify their identified constructs via confirmatory factor analysis (CFA) or exploratory structural equation modeling (ESEM).

Exploratory factor analysis, while an important initial step for assessing dimensionality, cannot provide the same level of information and clarity regarding factor structure as confirmatory factor analysis or modern equivalents such as exploratory structural equation modeling (Brown & Moore, 2012; Fabrigar et al., 1999; Flake et al., 2017). Confirmatory factor analysis, by contrast, is a theory-driven approach to assessing dimensionality that draws on structural equation modeling to allow testing of different model configurations and evaluation of model fit indices (Alamer, 2022; Woodrow, 2014). Unlike EFA, CFA can model the nature of relationships among items, account for measurement error, and allow for testing of hypothetical latent constructs and competing models. For these reasons, EFA is considered to be merely a first step in scale validation (Brown & Moore, 2012). To the authors' knowledge, no research has yet investigated the dimensionality of the STABLE-2007 while applying best practices in factor analysis.

Other researchers have proposed alternative, hypothetical factor solutions for the STABLE-2007. Although they did not directly examine construct validity via exploratory or confirmatory factor analysis, Brankley et al. (2021, p. 51) argue that the STABLE-2007's items assessing sexual recidivism risk may reflect several uniquely predictive aspects of sexual offending (e.g., risk for pedophilia vs. sexual sadism). This argument, notably, is consistent with the Structured Risk Assessment (SRA) Framework, a theoretical framework for dynamic risk assessment of men convicted of sexual offense (Thornton, 2002, 2013). The SRA framework includes four criminogenic needs domains: sexual interests, distorted attitudes, relational style, and self-management. In addition, each domain within the SRA framework is theoretically related to between two and three

subdomains (e.g., the *self-management* domain consists of two sub-domains reflecting *social deviance* and *dysfunctional coping in response to stress/problems*). This framework, in other words, implies that the STABLE-2007 should assess at least four dimensions related to sexual recidivism. Yet, with the exception of recent network analyses applied to STABLE-2007 scores in Canada, Alaska, and Iowa, the interrelationships among these domains remain unexplored (Heffernan & Ward, 2017; van den Berg et al., 2020, 2022).

Beyond determining the dimensional structure of an instrument's scores, demonstrating strong construct validity also requires assessment of convergent and discriminant validity (Strauss & Smith, 2009). Convergent validity refers to the degree that two measures that should be theoretically related are related. Conversely, discriminant validity refers to the degree that measures of constructs theoretically unrelated are in fact unrelated. Evidence regarding the STABLE-2007's convergent and discriminant validity is limited and mixed. Smeth (2013) examined total score correlations between the STABLE-2007, Static-99R, and Acute-2007, determining that the STABLE-2007 assesses different constructs than static ($r = 0.28$) and acute-dynamic ($r = 0.30$) risk. Nunes and Babchishin (2012) investigated the construct validity of STABLE-2007 by examining correlations between individual items and factors from other related scales (e.g., the RAPE and MOLEST scales and the UCLA Loneliness Scale, Bumby, 1996; Russell, 1996), concluding that STABLE items may be measuring different underlying constructs than those suggested by the original item names. Neither Smeth (2013) nor Nunes and Babchishin (2012), however, conducted formal analyses of the STABLE-2007's factor structure. Instead, the authors relied on analysis of total scores or individual item scores, limiting the conclusions that could be drawn regarding the construct validity of latent risk factors within the STABLE-2007.

Measurement Invariance: Assessing Reliability as a Prerequisite for Validity

Measurement invariance refers to the idea that the dimensional structure of a psychometric instrument should be consistent across time and contexts (e.g., cross-cultural comparisons; see Hagger et al., 2003; Megreya et al., 2016). Measurement invariance analysis, specifically, is an extension of confirmatory factor analysis that uses structural equation modeling to determine whether the factor structure of a psychometric instrument is stable across time and context (e.g., Megreya et al., 2016). Assessing invariance in the factor structure of psychometric instruments is a critical component of reliability in psychological assessment, as an unstable factor structure over time can be an indicator of low reliability in the instrument. Reliability is a necessary, yet insufficient, prerequisite for validity. Without consistency in measurement, one cannot conclude that an instrument is valid, even if a scale is shown to otherwise produce excellent predictive validity (Clifton, 2020; Cook & Beckman, 2006). By assessing measurement invariance over time, therefore, this study provides the necessary foundation for establishing that

the STABLE-2007 is not only valid but reliable as a risk assessment instrument in the United States and elsewhere.

Method

Participants

All analyses were conducted using data from an existing database of 873 adult male parolees who received between two and four STABLE-2007 ratings while undergoing community supervision. The average time between the first and fourth STABLE-2007 assessment ranged from 1 to 4 years ($Mdn = 3.00$, $M = 2.70$, $SD = 0.66$). All participants were rated on the STABLE-2007 at least twice, 488 participants were rated on the STABLE-2007 at least three times, and 171 participants were rated at least four times. Given that fewer than 5% of participants received five or more STABLE-2007 ratings ($n = 41$, 4.70%), we restricted our analyses to only those participants who completed between two and four assessments. Demographic and descriptive statistics for our sample are provided in Table 1. This study was approved by the CDCR Research Oversight Committee, the California State University, Fresno Institutional Review Board (IRB), and the California Committee for the Protection of Human Subjects, which serves as the IRB for all California state agencies.

Treatment Program Description

Sex offender treatment programs in California for individuals who are registered and under probation or parole supervision are required to meet standards established by the California Sex Offender Management Board (CASOMB). These standards are grounded in the Risk, Need, and Responsivity (RNR) principles that guide assessment and treatment of individuals convicted of sexual offenses. Assessment includes the Static-99R to establish baseline risk and inform the intensity of treatment, with the STABLE-2007 and Level of Service/Case Management Inventory (LS/CMI) administered at intake and annually to monitor changes in risk and treatment needs. Each agency has internal procedures for storing and using assessment data. In addition to standardized protocols for treatment completion, maintenance, and aftercare, agencies are also required to report assessment results, including individual item scores, through a government-maintained automated reentry management system. The data used in the present study were obtained from this government reporting system, which is commonly referred to as the Automated Reentry Management System (ARMS). Treatment is guided by a combination of internally developed and commercially available manuals, together with cognitive-behavioral and social learning approaches designed to address dynamic risk factors.

Table 1
Demographic and Descriptive Information (N = 873)

Variables	M (SD)	% (valid n)
Demographic information		
Age at program entry	44.03 (13.60)	99.5% (869)
Ethnicity		99.4% (868)
African American		21.6% (189)
Asian		2.4% (21)
American Indian		2.1% (18)
Hispanic		38.1% (333)
White		30.9% (270)
Other		4.2% (37)
Risk scales		
STATIC-99R Total Score	2.93 (2.62)	97.9% (855)
STABLE-2007 Total Score ^a	5.96 (3.79)	100.0% (873)
Time I	6.60 (4.44)	100.0% (873)
Time II	5.63 (4.15)	100.0% (873)
Time III	5.19 (3.78)	55.9% (488)
Time IV	5.37 (3.90)	19.6% (171)

Note. Subsample size (n) fluctuates as a function of missing data.
^aFewer than 5% of respondents provided scores on assessment time points V through VII (n = 41, 4.70%), so we restricted our factor analyses to assessment time points I through IV. The grand mean for STABLE-2007 scores (M = 5.96, SD = 3.79), however, was calculated based on all available data.

Measures

STABLE-2007

The STABLE-2007 is a risk assessment instrument designed to assess 13 criminogenic needs among men convicted of sexual offenses (Hanson et al., 2007). It includes the following items: (1) Significant Social Influences, (2) Capacity for Relationship Stability, (3) Emotional Identification with Children, (4) Hostility Toward Women, (5) General Social Rejection/Loneliness, (6) Lack of Concern for Others, (7) Impulsive Acts, (8) Poor Cognitive Problem Solving, (9) Negative Emotionality, (10) Sex Drive and Preoccupation, (11) Sex as Coping, (12) Deviant Sexual Interests, and (13) Cooperation with Supervision. Each item is scored on a 3-point-rating scale coded as 0 for “no problem,” 1 for “some concern/slight problem,” and 2 for “present/definite concern.” Items are evaluated independently via a semi-structured interview, review of the individual’s criminal history, and by considering an individual’s prospective functioning presuming conditions of

freedom. Assessed individuals are assigned a total score ranging between 0 and 26, with higher scores indicating a higher risk of recidivism. Meta-analytic reviews of previous research (Brankley et al., 2021) indicate that the Stable-2007 instrument produces scores with strong interrater reliability (median ICC = .90) and moderate predictive validity (AUC = .674, 95% CI [.646, .702]).

Static-99R

The Static-99R is an actuarial risk assessment instrument designed to assess 10 static risk factors among men convicted of sexual offenses (Helmus et al., 2022; Phenix et al., 2017). It includes the following items: (1) Age at release, (2) Any live-in intimate relationship for 2 or more years, (3) Convictions for index non-sexual violence, (4) Prior convictions for non-sexual violence, (5) Prior sex offenses, (6) Four or more prior sentencing dates, (7) Convictions for noncontact sexual offenses, (8) Any unrelated victims, (9) Any stranger victims, and (10) Any male victims. Items are evaluated via a review of the individual's demographic and criminal history, with total scores ranging from -3 to 12. With the exception of Items 1 and 5, items are scored as either 0 (*not present*) or 1 (*present*). Item 1 is scored on a scale from -3 (*age 60 or older*) to 1 (*age 18 to 34.9*). Item 5 is scored on a scale from 0 (*no charges or convictions*) to 3 (*six or more charges or 4 or more convictions*). Only total scores were provided in this dataset. Similar to the STABLE-2007, meta-analytic reviews indicate that the Static-99R instrument produces scores with strong interrater reliability (ICCs ranging between .84 and .95) and moderate predictive validity (Helmus et al., 2022; fixed effect AUC = .678, 95% CI [.669, .668]; random effects AUC = .689, 95% CI [.671, .706]).

Data Analysis

Exploratory Structural Equation Modeling

To assess the underlying dimensions of STABLE-2007 scores within our sample, we conducted exploratory structural equation modeling (ESEM) for item scores at each of the first four assessment time points (Asparouhov & Muthén, 2009; Marsh et al., 2014). Exploratory structural equation modeling, which applies an EFA measurement model with rotations within a structural equation modeling framework, integrates the rigor and parsimony of Confirmatory Factor Analysis (CFA) with the flexibility of Exploratory Factor Analysis (EFA) (for a detailed review of ESEM, see Marsh et al., 2014). The advantage of ESEM over CFA is that it allows for items to cross-load with other factors, whereas CFA forces cross-loadings to be 0. ESEM thereby addresses several issues with traditional CFA approaches, in that even small cross-loadings (<.100) can bias parameter estimates and it is extraordinarily rare for psychological constructs to be completely independent from one another (Asparouhov et al., 2015; Morin et al., 2016).

We constructed our ESEM models via the *psych* and *lavaan* packages in R (Revelle, 2025; Rosseel, 2012). Given the ordinal structure of STABLE-2007 items, we assumed

a polychoric correlation structure and therefore extracted factors using mean-and-variance-adjusted (WLSMV) estimation (Goretzko et al., 2021; Holgado-Tello et al., 2010); extracted factors were rotated obliquely via the Oblimin method. The number of factors retained was based on parallel analysis (Horn, 1965), Velicer's MAP criterion (Velicer, 1976; Velicer et al., 2000), and comparative model fit across the first four assessment time points.

We evaluated the fit of our final model based on the following indices: (a) the mean-and-variance-adjusted chi-square statistic (Asparouhov & Muthén, 2021), (b) the robust root mean square error of approximation (R-RMSEA) with a value smaller than or equal to .05 indicating a good fit and a value smaller than or equal to .08 indicating acceptable fit (Savalei, 2018; Schermelleh-Engel et al., 2003), (c) the standardized root mean residual (SRMR) with a value smaller than or equal to .08 indicating good fit (Hu & Bentler, 1999), and (d) the robust comparative fit index (R-CFI) with a value of .97 or higher indicating good fit and a value of .95 or higher indicating acceptable fit (Hu & Bentler, 1999; Savalei, 2018). Robust RMSEA and CFI values were used instead of traditional RMSEA and CFI values due to the STABLE-2007's ordinal data structure, as robust indices are calculated using the mean-and-variance-adjusted chi-square statistic (Asparouhov & Muthén, 2021; Pavlov et al., 2020). Likewise, we report ordinal alpha coefficients (ordinal α) to assess each identified factor's internal consistency via the *misty* package in R (Yanagida, 2025; Zumbo & Kroc, 2019; Zumbo et al., 2007).

Measurement Invariance

We assessed measurement invariance over time for the two-factor model using multiple-group confirmatory factor analysis (MG-CFA) via the *semTools* and *Lavaan* packages in R (Jorgenson et al., 2026; Rosseel, 2012). Given the ordinal structure of STABLE-2007, we followed guidelines for assessing invariance with categorical outcomes outlined by Wu and Estabrook (2016) and Svetina et al. (2020). The first and least restrictive step is to test a *baseline* or *configural* variance model wherein the number, pattern, and thresholds for parameters are assumed to be equal across time points, but values of parameters are free to vary. The second step is to test a *metric* invariance model wherein the pattern and values of factor loadings are assumed to be equivalent across time points. The final step is to test a *scalar* invariance model wherein both loadings and intercepts are assumed to be equal across time points. We used Svetina and Rutkowski's (2017) recommendations regarding measurement invariance cutoffs for ordinal data: (1) $\Delta\text{RMSEA} \leq .05$ in conjunction with significant $\Delta\chi^2$ as evidence of *metric* invariance and (2) $\Delta\text{RMSEA} \leq .01$ and $\Delta\text{CFI} \leq -.002$ as evidence of *scalar* invariance.

Results

Exploratory Structural Equation Modeling

Parallel analysis and Velicer's MAP test suggested viable factor solutions ranging from 1 to 5 across all assessment time points. To select a final model for assessment of measurement invariance, we estimated and compared model fit for one-, two-, three-, four-, and five-factor solutions at each time point, the results of which are presented in Table 2. Although a four-factor model technically produced the strongest model fit statistics, both the two-factor ($R\text{-RMSEA} \leq .049$, $SRMR \leq .070$, $CFI \geq .963$) and three-factor ($R\text{-RMSEA} \leq .040$, $SRMR \leq .053$, $CFI \geq .983$) models met criteria for strong model fit across all time points. Therefore, following the principle of parsimony (Falk & Muthukrishna, 2023), we selected the two-factor solution as our final model for the purpose of measurement invariance evaluation. To aid future replication efforts, we provide a detailed overview of the two-, three-, and four-factor ESEM solutions below.

The two-factor solution, as shown in Figure 1 below, resulted in two highly interpretable constructs ($|\lambda|_{\text{Antisociality}} = 0.502\text{--}0.796$, $M = 0.661$, ordinal $\alpha = 0.86$; $|\lambda|_{\text{AtypicalSexuality}} = 0.491\text{--}0.795$, $M = 0.618$, ordinal $\alpha = 0.66$). The first factor consisted of eight items reflecting *antisociality*: (1) Significant Social Influences, (4) Hostility Toward Women, (5) General Social Rejection, (6) Lack of Concern for Others, (7) Impulsive Acts, (8) Poor Cognitive Problem Solving, (9) Negative Emotionality, and (13) Cooperation with Supervision. The second factor consisted of four items reflecting *atypical sexuality*: (3) Emotional Identification with Children, (10) Sex Drive and Preoccupation, (11) Sex as Coping, and (12) Deviant Sexual Interests. Item 2 (Capacity for Relationship Stability), however, did not load onto either factor ($\lambda_{\text{Antisociality}} = 0.264$; $\lambda_{\text{AtypicalSexuality}} = 0.182$). Our two-factor solution suggests that California adult male parolees' scores on the STABLE-2007 may reflect two sources of risk: antisocial tendencies and atypical sexuality.

The three-factor solution, similarly, resulted in *antisociality* and *atypical sexuality* constructs ($|\lambda|_{\text{Antisociality}} = 0.478\text{--}0.734$, $M = 0.649$, ordinal $\alpha = 0.86$; $|\lambda|_{\text{AtypicalSexuality}} = 0.432\text{--}0.762$, $M = 0.646$, ordinal $\alpha = 0.65$), but also identified a third *asocial relations* construct consisting of two items: general loneliness and emotional identification with children ($|\lambda|_{\text{AsocialRelations}} = 0.453\text{--}0.594$, $M = 0.524$, ordinal $\alpha = 0.46$). The four-factor solution, in addition to *atypical sexuality* ($|\lambda|_{\text{AsocialRelations}} = 0.421\text{--}0.774$, $M = 0.640$, ordinal $\alpha = 0.78$), split the *antisociality* construct into three smaller constructs representing *asocial relations* ($|\lambda|_{\text{AsocialRelations}} = 0.469\text{--}0.561$, $M = 0.515$, ordinal $\alpha = 0.46$), *negative emotionality* ($|\lambda|_{\text{NegativeEmotionality}} = 0.630\text{--}0.793$, $M = 0.716$, ordinal $\alpha = 0.80$), and *impulsive antisociality* ($|\lambda|_{\text{ImpulsiveAntisociality}} = 0.491\text{--}0.770$, $M = 0.644$, ordinal $\alpha = 0.65$). Item 2 (Capacity for Relationship Stability), once again, did not load onto any factor across the three- or four-factor solutions.

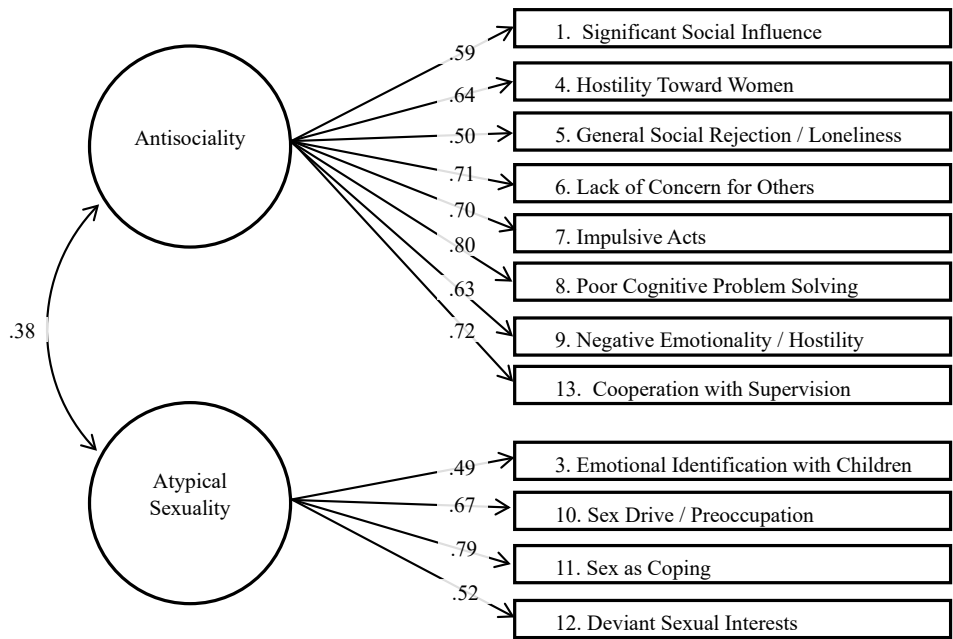
Table 2
Comparisons of Exploratory Structural Equation Model (ESEM) Fit Across Assessment Time Points

Model	χ^2 (df)	R-RMSEA	SRMR	R-CFI
Time I (n = 873)				
1-factor model	366.23 (65)	.060	.069	0.928
2-factor model	249.51 (53)	.049	.050	0.966
3-factor model	159.27 (42)	.040	.039	0.986
4-factor model	107.16 (32)	.035	.030	0.994
5-factor model ^a	272.57 (33)	.079	.058	0.808
Time II (n = 873)				
1-factor model	285.38 (65)	.054	.067	0.952
2-factor model	189.76 (53)	.041	.044	0.982
3-factor model	114.21 (42)	.031	.032	0.996
4-factor model	64.36 (32)	.023	.022	1.000
5-factor model	35.38 (23)	.016	.015	1.000
Time III (n = 488)				
1-factor model	159.01 (65)	.048	.058	0.945
2-factor model	123.86 (53)	.043	.046	0.964
3-factor model	68.80 (42)	.028	.033	0.988
4-factor model	35.20 (32)	.011	.021	0.999
5-factor model	20.19 (23)	.000	.015	1.000
Time IV (n = 171)				
1-factor model	93.54 (65)	.049	.082	0.941
2-factor model	73.51 (53)	.043	.070	0.963
3-factor model	52.76 (42)	.032	.053	0.983
4-factor model	31.91 (32)	.000	.032	1.000
5-factor model ^a	160.84 (33)	.149	.120	0.721
All Times (n = 2,405)				
1-factor model	781.58 (65)	.055	.064	.938
2-factor model	524.69 (53)	.044	.044	.969
3-factor model	307.81 (42)	.035	.035	.986
4-factor model	159.38 (32)	.027	.027	.995
5-factor model	76.96 (23)	.020	.014	.999

Note. χ^2 (df) = mean-and-variance corrected chi-square (degrees of freedom). R-RMSEA = robust root mean square error of approximation. SRMR = standardized root mean residual. R-CFI = Robust Comparative Fit Index.
^aModel could not be identified when assuming correlated (oblique) factors; re-estimated assuming independent (orthogonal) factors.

Figure 1

ESEM Analysis of STABLE-2007 Item Scores for All Time Points: Correlated Two-Factor Model



Note. ESEM = Exploratory Structural Equation Modeling. Item 2 (“Capacity for Relationship Stability”) did not load onto either factor ($\lambda_{Antisociality} = 0.26$; $\lambda_{AtypicalSexuality} = 0.18$).

Convergent and Discriminant Validity: Changeable Versus Static Risk

Next, we assessed for convergent and discriminant validity between our identified *antisociality* and *atypical sexuality* changeable risk factors and static risk factors by calculating Spearman’s rho correlations among STABLE-2007 scores combined across all four assessment time points and initial Static-99R scores. Antisociality scores exhibited weak convergence with static risk ($r_s = .375, p < .001$) and atypical sexuality scores ($r_s = .238, p < .001$). Atypical sexuality scores exhibited strong discriminant validity when compared to static risk scores ($r_s = -.028, p = .416$). But, whereas the *antisociality* factor exhibited good internal consistency (ordinal $\alpha = 0.86$), the *atypical sexuality* factor exhibited relatively weak internal reliability (ordinal $\alpha = 0.66$), suggesting that the items within the *atypical sexuality* factor might be assessing multiple, modestly related, yet distinct aspects of atypical sexuality.

Measurement Invariance

For conciseness, we provide a summary and comparison of model fit indices across measurement invariance models in Table 3. Overall, we observed evidence for measurement invariance across time in our sample. Specifically, scores passed the tests for both metric invariance ($\Delta\text{RMSEA} = -.003$; $\Delta\chi^2_{(30)} = 49.92$, $p = .013$) and scalar invariance ($\Delta\text{RMSEA} = -.003$; $\Delta\text{R-CFI} = .000$).

Table 3
Summary of Measurement Invariance Analyses for Two-Factor Model

Model	χ^2 (df)	R-RMSEA	SRMR	R-CFI
Model fit indices				
Configural model	765.37 (212)***	.066	.081	0.934
Metric model	815.29 (242)***	.063	.084	0.932
Scalar model	843.17 (272)***	.059	.088	0.932
Differences in model fit				
Configural vs. Metric	49.92 (30)*	-.003	.003	-0.002
Metric vs. Scalar	27.88 (30)	-.004	.004	0.000

Note. All models estimated using the two-factor solution identified via exploratory structural equation modeling. χ^2 (df) = mean-and-variance corrected chi-square (degrees of freedom). R-RMSEA = robust root mean square error of approximation. SRMR = standardized root mean residual. R-CFI = Robust Comparative Fit Index.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Discussion

Using a modern structural equation modeling approach, the present study examined the dimensional structure and measurement invariance properties of STABLE-2007 scores in a sample of male parolees who were mandated to participate in sexual offense treatment. Consistent with prior research on risk assessment for sexual offenses (Brouillette-Alarie & Hanson, 2015; Brouillette-Alarie et al., 2016), we identified a time-invariant two-factor solution. The first factor reflected an *antisociality* construct and the second factor reflected a *atypical sexuality* construct. These *antisociality* and *atypical sexuality* constructs were conceptually distinct from each other and from static risk measured via Static-99R scores. Multiple-group confirmatory factor analysis confirmed that this two-factor structure exceeded empirical standards for both metric invariance ($\Delta\text{RMSEA} = -.003$) and scalar invariance ($\Delta\text{RMSEA} = -.003$; $\Delta\text{R-CFI} = .000$), indicating that the *antisociality* and *atypical sexuality* constructs exhibit reliability with respect to dimensional structure over time (Svetina & Rutkowski, 2017). Notably, our two-, three, and four-factor solutions do not align with the informal five-factor model currently used in official STABLE-2007 training materials in California (Fernandez et al., 2014). Our findings suggest that these

training materials and the associated coding manual may need to be updated to reflect a more accurate representation of the changeable risk factors measured by the STABLE-2007.

Evaluating the Dimensionality of the STABLE-2007

By using a modern ESEM approach to assess factor structure and measurement invariance, our study provides the most statistically sophisticated examination of STABLE-2007's dimensionality to date. Whereas prior research has assessed the STABLE-2007's dimensionality via exploratory factor analysis or network analysis of individual items (Brouillette-Alarie & Hanson, 2015; Etzler et al., 2020; van den Berg et al., 2020, 2022), exploratory structural equation modeling (ESEM) enabled us to clarify the instrument's factor structure by integrating the rigor of Confirmatory Factor Analysis (CFA) with the flexibility of Exploratory Factor Analysis (Asparouhov & Muthén, 2009; Marsh et al., 2014). In doing so, our research establishes the STABLE-2007's validity and reliability in California and broader United States correctional contexts.

Although we selected a two-factor solution as our final model, it is worth noting that the one-, three-, and four-factor models exhibited good or acceptable model fit across the four assessment time points included within our study. Even the one-factor solution, which had the "worst" fit to the scores within our sample, would still be considered acceptable by our model selection standards, even if it did not exceed all model selection criteria. We selected the two-factor model as our final model because it exceeded criteria for "good" or "excellent" fit at every time point while being more parsimonious than the three- and four-factor models (Falk & Muthukrishna, 2023). Our two-factor solution also aligned more strongly with past theoretical (Brouillette-Alarie et al., 2016; Olver et al., 2007; McGrath & Thompson, 2012) and empirical (Brien-Robidoux, 2022; Etzler et al., 2020; Hanson et al., 2007) solutions than the three- and four-factor solutions. We note, however, that the distinction between our two-factor solution within California correctional populations (*antisociality* & *atypical sexuality*) and the three-factor solution within Austrian correctional populations identified by Etzler and colleagues (2020; *antisociality*, *sexual deviance*, & *hypersexuality*) suggests a need for further cross-cultural validation of the STABLE-2007 instrument. That is, future research should investigate whether the dimensional structure of STABLE-2007 is susceptible to change across cultural contexts.

Our findings also underline the complexity of dynamic risk assessment constructs related to sexual offending. Whereas our first *antisociality* factor exhibited good internal consistency (ordinal $\alpha = 0.86$), our second *atypical sexuality* factor exhibited moderately weak internal consistency (ordinal $\alpha = 0.66$). The relatively low reliability of the *atypical sexuality* construct suggests that the items within Factor 2 may be assessing multiple, moderately related, yet distinct aspects of atypical sexuality. Notably, this finding is consistent with Brankley et al.'s (2021) argument that the STABLE-2007 items may be assessing several underlying dimensions of sexual criminality. Moreover, the low internal

consistency for the *atypical sexuality* construct is consistent with broader theoretical frameworks for dynamic risk assessment of men convicted of sexual offense, most notably the Structured Risk Assessment (SRA) Framework (Thornton, 2002, 2013). The idea that atypical sexuality can be split into multiple sub-domains also aligns with recent, innovative approaches to measuring dimensionality via structured network analysis (van den Berg et al., 2020, 2022). Van den Berg et al.'s (2020) analysis of connections within the sexual-regulation-cluster, for example, mirrors Etzler et al.'s (2020) finding that *sexual preoccupation* and *sex as coping* form a factor conceptually separate from *deviant sexual interests*. We note, also, that the strong centrality of *general rejection/loneliness*, *poor cognitive problem solving*, and *impulsive acts* in van den Berg et al.'s (2020) data align with the high internal consistency for the *antisociality* factor observed in our sample scores.

Further supporting the idea that items related to relationships may be a unique construct within the STABLE-2007, the item *capacity for relationship stability* did not load onto any of the viable factor solutions identified within in our sample – even after combining the sample across all assessment time points to maximize statistical power. One possible explanation for this unexpected item loading pattern is that the *capacity for relationship stability* item was measured in a substantively different manner than the other items on the STABLE-2007 or reflects a distinct underlying construct. This is not to suggest, however, that the *capacity for relationship stability* item should be removed from the STABLE-2007 instrument. Instead, we highlight this item loading pattern to provide guidance on future STABLE-2007 research investigating the instrument's predictive validity. In particular, researchers might consider testing whether *capacity for relationship stability* uniquely predicts recidivism risk above and beyond the *antisociality* and *atypical sexuality* constructs identified here. Prior research has already shown that STABLE-2007 items vary with respect to their potential for predicting reoffending. In van den Berg et al.'s (2020) shortest paths analysis, for example, *impulsive acts* shared the strongest independent relationship with sexual and violent recidivism, whereas *emotional identification with children* was directly related to sexual, but not violent, recidivism. In the event that *capacity for relationship stability* lacks unique predictive validity, researchers might next consider testing whether including or excluding this item from models improves the predictive validity of STABLE-2007 scores. Alternatively, in the event that this item retains unique predictive validity, researchers might consider creating and testing additional items that align with the underlying construct measured by the *capacity for relationship stability* item.

Our results also suggest that the *antisociality* construct could be represented by two or more specific subdomains. Specifically, our four-factor solution, consistent with recent research on Dark Tetrad personality traits and atypical sexuality, suggests that the broader *antisociality* domain could be feasibly split into two dimensions representing *negative emotionality* and *impulsive antisociality* (Costa et al., 2023; Lassche et al., 2024). Our three- and four-factor ESEM solutions also provide moderate evidence that general

loneliness and emotional identification with children can be treated as an independent *asocial relations* construct that is distinct from both *atypical sexuality* and *antisociality* – though the reliability of this construct was low (ordinal $\alpha = 0.46$). This finding aligns with prior research indicating that individuals with pedophilia tend to report higher levels of loneliness and social anxiety (Marshall et al., 2012; Schulz et al., 2017). We recommend that future research investigate whether these more complex solutions provide incremental predictive validity above and beyond our two-factor solution, as the ability of unique constructs to predict recidivism is highly useful in risk assessment (Mann et al., 2010).

Finally, we found strong support for the STABLE-2007's ability to reliably measure our identified *antisociality* and *atypical sexuality* constructs over the course of community supervision. Specifically, scalar invariance in the dimensional structure of STABLE-2007 scores over time suggests that, within California, the two-factor solution for the STABLE-2007 benefits from clear structure and long-term construct validity, as previously observed in other contexts (i.e., Etzler et al., 2020). We emphasize, however, that these findings should be interpreted within the context of California parolees. Reliability and validity are properties of scores within a given sample, not properties of an instrument (Cook & Beckman, 2006). In other words, the validity and reliability characteristics identified within this study may not extend to correctional and treatment contexts beyond California specifically and the United States more generally. With this caveat in mind, the present findings provide strong preliminary evidence that the *antisociality* and *atypical sexuality* constructs within the STABLE-2007 can be measured reliably over time.

Future Directions and Limitations

The present findings have several implications for treatment and evaluation of men convicted of sex offenses. Consistent with Smeth (2013), we observed moderate evidence for discriminant validity between the STABLE-2007's *antisociality* and *atypical sexuality* dynamic risk constructs and static risk as measured by Static-99R total scores. These findings suggest that the antisociality and atypical sexuality factors identified in this study are distinct from one another and from static risk factors assessed by instruments such as the Static-99R. We could not, however, examine discriminant validity by comparing against Static-99R subscales in this study, as the available data did not include Static-99R item scores; we therefore recommend that future research examine the discriminant validity of instrument subscale scores, if possible.

Moreover, future research will need to verify the construct validity of our identified *antisociality* and *atypical sexuality* constructs by evaluating convergence with other scales assessing these constructs. Nunes and Babchishin (2012) previously observed low correlations between individual STABLE-2007 items and purportedly related scales such as the RAPE and MOLEST scales and the UCLA Loneliness Scale (Bumby, 1996; Russell,

1996). Assessing convergence using *antisociality* and *atypical sexuality* subscale scores rather than individual item scores might provide a clearer picture of the instrument's convergent and discriminant validity with respect to relevant constructs such as impulsivity, loneliness, empathy, and sexual preoccupation. However, we acknowledge that restrictions on data collection within California correctional facilities may hamper or prevent administration of non-state-mandated assessment tools. Therefore, although imperfect, we recommend first assessing convergence with general criminality via other state-mandated instruments such as the LS/CMI (Jimenez et al., 2018; Schmidt et al., 2017).

Next, we highlight an important limitation for our approach to measurement invariance in this paper, namely that Svetina et al.'s (2020) multiple-group CFA framework cannot account for the assumption of dependence inherent to longitudinal designs. Although longitudinal measurement invariance models exist to address this issue (e.g., Liu et al., 2017; Mackinnon et al., 2022), these models were unsatisfactory solutions for the current data for several reasons. First, Liu et al.'s (2017) framework, while adequate at handling ordinal data structures, was developed for assessing unidimensional scales, not multidimensional scales like the STABLE-2007. Second, traditional longitudinal measurement invariance frameworks like the one developed Mackinnon et al. (2022) cannot account for the ordinal structure of the STABLE-2007, resulting in ill-fitting or unidentified models. At this time, Svetina et al.'s (2020) approach is the only measurement invariance approach, to the authors' knowledge, that can both handle multidimensional scales and account for ordinal data structures. Given that the core finding of this paper is that the STABLE-2007 is best captured by two or more factors and that capturing these dimensions necessitates accounting for the instrument's ordinal data structure, we determined that Svetina et al.'s (2020) multiple-group CFA framework offered a decent, albeit imperfect, compromise for assessing the reliability of the STABLE-2007's factor structure over time. However, we highlight this issue as a critical area for future re-analysis once longitudinal measurement invariance models that account for multidimensional ordinal data are developed.

Finally, and perhaps most importantly, validation of the STABLE-2007 lays the foundation for future research examining the predictive validity of STABLE-2007 scores. There is still a strong need for additional evaluation of construct and predictive validity in risk assessment instruments, as several risk assessment instruments commonly used in California have been shown to have poor or moderate predictive validity (Fazel et al., 2022). The STABLE-2007 has previously been found to have moderate accuracy in predicting sexual recidivism (e.g., Brankley et al., 2021). Assessing the predictive validity of construct-specific scores rather than relying on the raw STABLE-2007 total scores, however, may improve the STABLE-2007's ability to reliably predict general and sexual reoffending.

A construct-focused approach would also enable testing of whether *antisociality* or *atypical sexuality* is a stronger predictor of recidivism risk. Prior research on static risk suggests that the inclusion of sex-crime-specific items can undermine an instrument's ability to predict general and violent recidivism (Babchishin et al., 2016). As noted by Brankley et al. (2021), this suggests that the STABLE-2007 *atypical sexuality* construct may be unrelated, or even negatively related, to nonsexual reoffending. If this hypothesis is supported, practitioners using the STABLE-2007 who are interested in predicting general or violent recidivism might benefit from stronger predictive validity if they rely on *antisociality* subscale scores rather than total scores.

Conclusion

We observed moderate initial support for the validity and reliability of the STABLE-2007 risk assessment instrument within a sample of California parolees under community supervision. Consistent with prior theory regarding assessment of sexual recidivism risk, we identified two distinct time-invariant constructs in STABLE-2007 scores: *antisociality* and *atypical sexuality*. We recommend that future research build upon the established predictive validity of STABLE-2007 scores by examining the relative diagnostic utility of the *antisociality* and *atypical sexuality* constructs identified in this article within other United States correctional populations. We also recommend that training materials be updated to reflect the empirically verified constructs identified in this research, as accurate representation of risk factors may improve treatment and evaluation of men convicted of sex offenses.

Funding: This research was supported in part through grant number ROC#2107-053-ROC.

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

Competing Interests: The authors have declared that no competing interests exist.

Author Note: The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views or opinions of the California Department of Corrections and Rehabilitations (CDCR), including the Division of Adult Parole Operations (DAPO), Research Oversight Committee or California Sex Offender Management Board.

Data Availability: The dataset used in this study contains potentially identifying and sensitive information and is not publicly available due to data privacy laws. Access to these data can be requested by contacting the California Department of Corrections and Rehabilitation (CDCR) Office of Research (email: data.requests@cdcr.ca.gov).

Supplementary Materials

Although we focused our results and discussion on the 2-factor ESEM model in this paper, we recognize that researchers may be interested in alternative models for the purpose of future replication and validation efforts. Accordingly, we have provided a supplementary table containing the factor loadings for the unselected 1-, 3-, and 4-factor ESEM models via the open-access repository PsychArchives (for access, see [Ditchfield et al., 2026S](https://doi.org/10.23668/psycharchives.22440)).

Index of Supplementary Materials

Ditchfield, R. E., Azizian, A., D'Orazio, D., & Chankin, L. (2026S). *Supplementary materials to "Assessing the dimensional structure and measurement invariance of STABLE-2007 scores in a California sample of adult men on parole for sex crimes"* [Supplementary table]. PsychOpen GOLD. <https://doi.org/10.23668/psycharchives.22440>

References

- Alamer, A. (2022). Exploratory structural equation modeling (ESEM) and bifactor ESEM for construct validation purposes: Guidelines and applied example. *Research Methods in Applied Linguistics*, 1(1), Article 100005. <https://doi.org/10.1016/j.rmal.2022.100005>
- Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16(3), 397–438. <https://doi.org/10.1080/10705510903008204>
- Asparouhov, T., & Muthén, B. (2021). Robust chi-square in extreme and boundary conditions: Comments on Jak et al. *Psych*, 3(3), 542–551. <https://doi.org/10.3390/psych3030035>
- Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeier et al. *Journal of Management*, 41(6), 1561–1577. <https://doi.org/10.1177/0149206315591075>
- Babchishin, K. M., Hanson, R. K., & Blais, J. (2016). Less is more: Using Static-2002R subscales to predict violent and general recidivism among sexual offenders. *Sexual Abuse: A Journal of Research and Treatment*, 28(3), 187–217. <https://doi.org/10.1177/1079063215569544>
- Beggs, S. M., & Grace, R. C. (2011). Treatment gain for sexual offenders against children predicts reduced recidivism: A comparative validity study. *Journal of Consulting and Clinical Psychology*, 79(2), 182–192. <https://doi.org/10.1037/a0022900>
- Bonta, J., & Andrews, D. A. (2023). *The psychology of criminal conduct* (7th ed.). Routledge.
- 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: A Journal of Research and Treatment*, 33(1), 34–62. <https://doi.org/10.1177/1079063219871572>
- Brien-Robidoux, E. (2022). *Dynamic risk assessment of sexual offenders in the real world: Study of predictive validity and dimensionality of the Static-99R and Stable-2007 on a French-speaking Canadian sample* [Master's thesis, Université de Montréal]. Papyrus. <https://doi.org/10.71781/23674>

- 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>
- Brouillette-Alarie, S., & Hanson, R. K. (2015). Comparaison de deux mesures d'évaluation du risque de récidive des délinquants sexuels [Comparison of two measures of recidivism risk assessment of sexual offenders]. *Canadian Journal of Behavioural Science / Revue Canadienne des Sciences du Comportement*, 47(4), 292–304. <https://doi.org/10.1037/cbs0000019>
- Brown, T. A., & Moore, M. T. (2012). Confirmatory factor analysis. In R. H. Hoyle (Ed.), *Handbook of structural equation modeling* (pp. 361–379). The Guilford Press.
- Bumby, K. M. (1996). Assessing the cognitive distortions of child molesters and rapists: Development and validation of the MOLEST and RAPE scales. *Sexual Abuse: A Journal of Research and Treatment*, 8(1), 37–54. <https://doi.org/10.1177/107906329600800105>
- Clifton, J. D. W. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods*, 25(3), 259–270. <https://doi.org/10.1037/met0000236>
- Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: Theory and application. *The American Journal of Medicine*, 119(2), 166.e7–166.e16. <https://doi.org/10.1016/j.amjmed.2005.10.036>
- Costa, R., Fávero, M., Moreira, D., Del Campo, A., & Sousa-Gomes, V. (2023). Dark Tetrad, acceptance of sexual violence, and sexism. *European Psychologist*, 28(1), 12–23. <https://doi.org/10.1027/1016-9040/a000485>
- Eisenberg, M. J., van Horn, J. E., Dekker, J. M., Assink, M., van der Put, C. E., Hendriks, J., & Stams, G. J. J. M. (2019). Static and dynamic predictors of general and violent criminal offense recidivism in the forensic outpatient population: A meta-analysis. *Criminal Justice and Behavior*, 46(5), 732–750. <https://doi.org/10.1177/0093854819826109>
- 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>
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. <https://doi.org/10.1037/1082-989X.4.3.272>
- Falk, C. F., & Muthukrishna, M. (2023). Parsimony in model selection: Tools for assessing fit propensity. *Psychological Methods*, 28(1), 123–136. <https://doi.org/10.1037/met0000422>
- Fazel, S., Burghart, M., Fanshawe, T., Gil, S. D., Monahan, J., & Yu, R. (2022). The predictive performance of criminal risk assessment tools used at sentencing: Systematic review of validation studies. *Journal of Criminal Justice*, 81, Article 101902. <https://doi.org/10.1016/j.jcrimjus.2022.101902>
- Fernandez, Y., Harris, A. J. R., Hanson, R. K., & Sparks, J. (2014). *STABLE-2007 coding manual: Revised 2014*. Public Safety Canada.

- Flake, J. K., Pek, J., & Hehman, E. (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological & Personality Science*, 8(4), 370–378. <https://doi.org/10.1177/1948550617693063>
- Goretzko, D., Pham, T. T. H., & Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510–3521. <https://doi.org/10.1007/s12144-019-00300-2>
- Hagger, M. S., Biddle, S. J. H., Chow, E. W., Stambulova, N., & Kavussanu, M. (2003). Physical self-perceptions in adolescence: Generalizability of a hierarchical multidimensional model across three cultures. *Journal of Cross-Cultural Psychology*, 34(6), 611–628. <https://doi.org/10.1177/0022022103255437>
- Hanson, R. K., & Harris, A. J. R. (1998). *Dynamic predictors of sexual recidivism* (User Report 1998-01). Ottawa, Ontario: Department of the Solicitor General of Canada. <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/dnmc-prdctrs-sxl/index-en.aspx>
- Hanson, R. K., & Harris, A. J. R. (2000). Where should we intervene? Dynamic predictors of sexual offence recidivism. *Criminal Justice and Behavior*, 27(1), 6–35. <https://doi.org/10.1177/0093854800027001002>
- 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*. Research Division, Public Safety Canada. <https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/ssssng-rsk-sxl-ffndrs/index-en.aspx>
- Heffernan, R., & Ward, T. (2017). A comprehensive theory of dynamic risk and protective factors. *Aggression and Violent Behavior*, 37, 129–141. <https://doi.org/10.1016/j.avb.2017.10.003>
- Helmus, L. M., Kelley, S. M., Frazier, A., Fernandez, Y. M., Lee, S. C., Rettenberger, M., & Boccaccini, M. T. (2022). Static-99R: Strengths, limitations, predictive accuracy meta-analysis, and legal admissibility review. *Psychology, Public Policy, and Law*, 28(3), 307–331. <https://doi.org/10.1037/law0000351>
- Holgado-Tello, F. P., Chacón-Moscoso, S., Barbero-García, I., & Vila-Abad, E. (2010). Polychoric versus Pearson correlations in exploratory and confirmatory factor analysis of ordinal variables. *Quality & Quantity*, 44(1), 153–166. <https://doi.org/10.1007/s11135-008-9190-y>
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jimenez, A. C., Delgado, R. H., Vardsveen, T. C., & Wiener, R. L. (2018). Validation and application of the LS/CMI in Nebraska probation. *Criminal Justice and Behavior*, 45(6), 863–884. <https://doi.org/10.1177/0093854818763231>
- Jorgenson, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2026). *semTools: Useful tools for structural equation modeling* (Version 0.5–9)[R]. [Computer software]. <https://CRAN.R-project.org/package=semTools>

- Lassche, M. M., Lasogga, L., de Roos, M. S., Leeflang, A., Ajazi, V., Axioti, M., Rassin, E., & van Dongen, J. D. M. (2024). Dark Tetrad personality traits, paraphilic interests, and the role of impulsivity: An EEG-study using a Go/No-Go paradigm. *Scientific Reports*, 14(1), Article 10884. <https://doi.org/10.1038/s41598-024-61214-z>
- Liu, Y., Millsap, R. E., West, S. G., Tein, J.-Y., Tanaka, R., & Grimm, K. J. (2017). Testing measurement invariance in longitudinal data with ordered-categorical measures. *Psychological Methods*, 22(3), 486–506. <https://doi.org/10.1037/met0000075>
- Mackinnon, S., Curtis, R., & O'Connor, R. (2022). A tutorial in longitudinal measurement invariance and cross-lagged panel models using *lavaan*. *Meta-Psychology*, 6. <https://doi.org/10.15626/MP.2020.2595>
- 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(2), 191–217. <https://doi.org/10.1177/1079063210366039>
- Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, 10, 85–110. <https://doi.org/10.1146/annurev-clinpsy-032813-153700>
- Marshall, L. E., O'Brien, M. D., Marshall, W. L., Booth, B., & Davis, A. (2012). Obsessive-compulsive disorder, social phobia, and loneliness in incarcerated internet child pornography offenders. *Sexual Addiction & Compulsivity*, 19(1–2), 41–52. <https://doi.org/10.1080/10720162.2012.665291>
- McGrath, A., & Thompson, A. P. (2012). The relative predictive validity of the static and dynamic domain scores in risk-need assessment of juvenile offenders. *Criminal Justice and Behavior*, 39(3), 250–263. <https://doi.org/10.1177/0093854811431917>
- Megreya, A. M., Latzman, R. D., Al-Attayah, A. A., & Alrashidi, M. (2016). The robustness of the nine-factor structure of the Cognitive Emotion Regulation Questionnaire across four Arabic-speaking Middle Eastern countries. *Journal of Cross-Cultural Psychology*, 47(6), 875–890. <https://doi.org/10.1177/0022022116644785>
- Miner, M. H., Brouillette-Alarie, S., & Hanson, R. K. (2023). Reliability and factor structure of the sex offender treatment intervention and progress scale. *Sexual Abuse: A Journal of Research and Treatment*, 35(1), 83–102. <https://doi.org/10.1177/10790632221082667>
- Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, 23(1), 116–139. <https://doi.org/10.1080/10705511.2014.961800>
- Nunes, K. L., & Babchishin, K. M. (2012). Construct validity of STABLE-2000 and STABLE-2007 scores. *Sexual Abuse: A Journal of Research and Treatment*, 24(1), 29–45. <https://doi.org/10.1177/1079063211404921>
- 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., Wong, S. C. P., Nicholaichuk, T., & Gordon, A. (2007). The validity and reliability of the Violence Risk Scale–Sexual Offender version: Assessing sex offender risk and evaluating therapeutic change. *Psychological Assessment*, 19(3), 318–329.
<https://doi.org/10.1037/1040-3590.19.3.318>
- Pavlov, G., Shi, D., & Maydeu-Olivares, A. (2020). Chi-square difference tests for comparing nested models: An evaluation with non-normal data. *Structural Equation Modeling*, 27(6), 908–917.
<https://doi.org/10.1080/10705511.2020.1717957>
- Phenix, A., Fernandez, Y., Harris, A., Helmus, M., Hanson, R. K., & Thornton, D. (2017). *Static-99R coding rules*. Research Division, Public Safety Canada.
<https://www.publicsafety.gc.ca/cnt/rsrscs/pblctns/sttc-2016/sttc-2016-en.pdf>
- Revelle, W. (2025). *psych: Procedures for psychological, psychometric, and personality research* (R package version 2.3.6) [Computer software]. <https://CRAN.R-project.org/package=psych>
- Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, 48(2), 1–36.
<https://CRAN.R-project.org/package=lavaanhttps://doi.org/10.18637/jss.v048.i02>
- Russell, D. W. (1996). UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. *Journal of Personality Assessment*, 66(1), 20–40. https://doi.org/10.1207/s15327752jpa6601_2
- Savalei, V. (2018). On the computation of the RMSEA and CFI from the mean-and-variance corrected test statistic with nonnormal data in SEM. *Multivariate Behavioral Research*, 53(3), 419–429. <https://doi.org/10.1080/00273171.2018.1455142>
- Schermelleh-Engel, K., Moosbrugger, H., & Muller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online*, 8(2), 23–74. <https://doi.org/10.23668/psycharchives.12784>
- Schmidt, N., Lien, E., Vaughan, M., & Huss, M. T. (2017). An examination of individual differences and factor structure on the LS/CMI: Does this popular risk assessment tool measure up? *Deviant Behavior*, 38(3), 306–317. <https://doi.org/10.1080/01639625.2016.1197003>
- Schulz, A., Bergen, E., Schuhmann, P., & Hoyer, J. (2017). Social anxiety and loneliness in adults who solicit minors online. *Sexual Abuse: A Journal of Research and Treatment*, 29(6), 519–540.
<https://doi.org/10.1177/1079063215612440>
- Seto, M. C., Augustyn, C., Roche, K. M., & Hilkes, G. (2023). Empirically-based dynamic risk and protective factors for sexual offending. *Clinical Psychology Review*, 106, Article 102355.
<https://doi.org/10.1016/j.cpr.2023.102355>
- Smeth, A. H. (2013). *Evaluating risk assessments among sex offenders: A comparative analysis of static and dynamic factors* [Master's thesis, Carleton University]. CURVE.
<https://doi.org/10.22215/etd/2013-09961>
- Strauss, M. E., & Smith, G. T. (2009). Construct validity: Advances in theory and methodology. *Annual Review of Clinical Psychology*, 5(1), 1–25.
<https://doi.org/10.1146/annurev.clinpsy.032408.153639>

- Svetina, D., & Rutkowski, L. (2017). Multidimensional measurement invariance in an international context: Fit measure performance with many groups. *Journal of Cross-Cultural Psychology*, 48(7), 991–1008. <https://doi.org/10.1177/0022022117717028>
- Svetina, D., Rutkowski, L., & Rutkowski, D. (2020). Multiple-group invariance with categorical outcomes using updated guidelines: An illustration using *Mplus* and the *lavaan/semTools* packages. *Structural Equation Modeling*, 27(1), 111–130. <https://doi.org/10.1080/10705511.2019.1602776>
- Thornton, D. (2002). Constructing and testing a framework for dynamic risk assessment. *Sexual Abuse: A Journal of Research and Treatment*, 14(2), 139–153. <https://doi.org/10.1177/107906320201400205>
- Thornton, D. (2013). Implications of our developing understanding of risk and protective factors in the treatment of adult male sexual offenders. *International Journal of Behavioral and Consultation Therapy*, 8(3–4), 62–65. <https://doi.org/10.1037/h0100985>
- van den Berg, J. W., Kossakowski, J. J., Smid, W., Babchishin, K. M., Borsboom, D., Janssen, E., van Beek, D., & Gijs, L. (2022). Dynamic risk factors in adult men who committed sexual offenses: Replication and comparison of networks found in two independent samples. *Psychology of Violence*, 12(6), 424–437. <https://doi.org/10.1037/vio0000445>
- van den Berg, J. W., Smid, W., Kossakowski, J. J., van Beek, D., Borsboom, D., Janssen, E., & Gijs, L. (2020). The application of network analysis to dynamic risk factors in adult male sex offenders. *Clinical Psychological Science*, 8(3), 539–554. <https://doi.org/10.1177/2167702620901720>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321–327. <https://doi.org/10.1007/BF02293557>
- Velicer, W. F., Eaton, C. A., & Fava, J. L. (2000). Construct explication through factor or component analysis: A review and evaluation of alternative procedures for determining the number of factors or components. In R. D. Goffin & E. Helmes (Eds.), *Problems and solutions in human assessment: Honoring Douglas N. Jackson at seventy* (pp. 41–71). https://doi.org/10.1007/978-1-4615-4397-8_3
- Wielinga, F. E., & Olver, M. E. (2024). Validity and reliability of the Violence Risk Scale–Sexual Offense version in a community sexual offense outpatient setting. *Psychological Assessment*, 36(6–7), 407–424. <https://doi.org/10.1037/pas0001318>
- Woodrow, L. (2014). Writing about factor analysis. In *Writing about quantitative research in applied linguistics* (pp. 110–121). https://doi.org/10.1057/9780230369955_10
- Wu, H., & Estabrook, R. (2016). Identification of confirmatory factor analysis models of different levels of invariance for ordered categorical outcomes. *Psychometrika*, 81(4), 1014–1045. <https://doi.org/10.1007/s11336-016-9506-0>
- Yanagida, T. (2025). *misty: Coefficient alpha and item statistics* (R package version 0.6.5) [Computer software]. <https://CRAN.R-project.org/package=misty>
- Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, 6(1), 21–29. <https://doi.org/10.22237/jmasm/1177992180>

Zumbo, B. D., & Kroc, E. (2019). A measurement is a choice and Stevens' scales of measurement do not help make it: A response to Chalmers. *Educational and Psychological Measurement*, 79(6), 1184–1197. <https://doi.org/10.1177/0013164419844305>

 International Association for the Treatment of Sexual Offenders <i>Sexual Offending: Theory, Research, and Prevention</i> (SOTRAP) is the official journal of the International Association for the Treatment of Sexual Offenders (IATSO).	 Leibniz-Institut für Psychologie PsychOpen GOLD is a publishing service provided by the Leibniz Institute for Psychology (ZPID), Germany.
---	--