1

Abstract: 209 words

Text: 3550 words

Table: 1

Figure: 1

Testing the reliability and construct validity of Short-Term Assessment of Risk and Treatability (START) in non-Western forensic psychiatric inpatients

Weerapong Sanmontree[a] and Peera Wongupparaj[b]

[a]Galya Rajanagarindra Institute, Department of Mental Health, Ministry of Public Health, Bangkok, Thailand

[b]Cognitive Science and Innovation Research Unit, College of Research Methodology and Cognitive Science, Burapha University, Thailand

Email addresses: drweerapongsanmontree@gmail.com; peera.wo@go.buu.ac.th

Correspondence to:

Peera Wongupparaj, BA, MA, PhD.

Cognitive Science and Innovation Research Unit, College of Research Methodology and Cognitive Science, Burapha University, Bang Saen, Thailand.

Fax: +66 (0)38 393 484, **Tel**: +66 (0) 38 102 078 ext. 130

Email: peera.wo@go.buu.ac.th

ORCID iDs

Weerapong Sanmontree   https://orcid.org/0000-0002-0969-6126

Peera Wongupparaj   https://orcid.org/0000-0001-8099-9157

Abstract

The Short-Term Assessment of Risk and Treatability (START) is deemed the most appropriate instrument for assessing violence risks and management because of its balanced approach between dynamic risk and protective factors. Although several facets of reliability and predictive validity of this tool were strong, its inter-rater reliability, construct validity, and implementation in Asian population were under-investigated. The objective of this research was to examine the inter-rater reliability and construct validity of the START: Thai version within forensic psychiatric inpatients. The participants consisted of 118 forensic psychiatric inpatients hospitalized at Galya Rajanagarindra Institute in Thailand. Trained mental health professionals (i.e., psychiatrists, forensic nurses, clinical psychologists, social workers, and occupational therapists) assessed each participant across twenty domains of the Thai START. The inter-rater reliability was examined using the intraclass correlation coefficient and a confirmatory factor analysis for ordinal data was used to test the construct validity of the scale. The main finding showed a good-to-excellent inter-rater reliability and supported two relational constructs (i.e., strength vs vulnerability subscales) of the Thai START. The Thai START is a promising tool for using in Thai forensic psychiatric setting but some items were not significant in contributing to the scale. This study also provides the guideline for implementing the tool in non-Western forensic psychiatric populations.

Key Words Thai START, Forensic psychiatric inpatients, Confirmatory factor analysis with ordinal data, Violence risk assessment.

1. Introduction

The Short-Term Assessment of Risk and Treatability (START) was developed in 2004 to address short-term or acute risk of violence in adolescents and forensic psychiatric settings (Nicholls, Brink, Desmarais, Webster, & Martin, 2006; Webster, Martin, Brink, Nicholls, & Middleton, 2004). This violence risk assessment has been widely used and regarded as a comprehensive and practical tool in clinical risk assessment alongside other standardized structured risk/forensic instruments (Cartwright, Desmarais, Hazel, Griffith, & Azizian, 2018; Hogan & Olver, 2018). According to a global survey across 44 countries on the use of measures to organize assessment, guide treatment plans and interventions, and monitor violence risk, the result from 2,135 mental health professionals (e.g., psychiatrist, psychologists, and nurses) demonstrated that the START was ranked fifth out of twelve instruments (Singh et al., 2014). Furthermore, a recent finding on the START implementation from forensic-clinical staff members finds the items to be easy to complete and facilitating communication among staff members (De Beuf, de Ruiter, & de Vogel, 2020).

Based on the START guideline, experienced raters are required to assess 20 dynamic items in terms of risk or 'vulnerabilities' and protective factors or 'strengths' (Webster, Martin, Brink, Nicholls, & Desmarais, 2009). The inclusion of protective factors was regarded as a critical advantage of this tool in that it may improve a positive perception of forensic psychiatric patients and reduce unethical and biased portrait of offenders (Braithwaite, Charette, Crocker, & Reyes, 2010; Rogers, 2000). The START includes social skills, relationships, occupational, recreational, self-care, mental state, emotional state, substance use, impulse control, external triggers, social

support, material resources, attitudes, medical adherence, rule adherence, conduct, insight, plans, coping, and treatability items.

The extant literature on the START psychometric properties suggested moderate-to-strong inter-rater reliability via Intraclass Correlation Coefficient (ICC) values (Abidin et al., 2013; Lowder, Desmarais, Rade, Johnson, & Van Dorn, 2019; S. Viljoen, Nicholls, Greaves, de Ruiter, & Brink, 2011; Whittington et al., 2014; Wilson, Desmarais, Nicholls, & Brink, 2010), moderate-to-good predictive validity (Braithwaite et al., 2010; Desmarais, Nicholls, Read, & Brink, 2010; Nicholls et al., 2006; Nonstad et al., 2010; Wilson et al., 2010), and moderate concurrent validity with other standardized risk assessment scales, the Historical-Clinical-Risk Management-20 (HCR-20) (Cartwright et al., 2018; Desmarais, Nicholls, Wilson, & Brink, 2012; Hogan & Olver, 2018), the Structured Assessment of Violence Risk in Youth (SAVRY) (J. L. Viljoen et al., 2012), the Structured Assessment of Protective factors (SAPROF) (Abidin et al., 2013), and the DUNDRUM Quartet (Abidin et al., 2013).

The construct validity was defined as an integration of empirical evidence and theoretical rationales that affects the interpretation or meaning of test scores (Messick, 1989; Strauss & Smith, 2009). Nonetheless, there has been no research on testing structural relationships between each item and its subscales and also between two subscales of the START at the latent factor level (McArdle, 1996). The significant structural relationships between two START subscales may suggest a construct validity (Atkinson et al., 2011; Tavakol & Wetzel, 2020), which it has so far not been considered and examined. Several researchers and clinicians have supported the increment validity of the START for incorporating the strength subscale into the risk assessment

framework (Braithwaite et al., 2010; Desmarais et al., 2012; O'Shea, Picchioni, & Dickens, 2016; Paetsch, van Os, Troquete, & van den Brink, 2019). The concept of incorporating the protective factor or the START strength subscale is debated at a fundamental level in terms of whether it is a unique entity or is simply the reverse of a risk factor or START vulnerability subscale (Haines et al., 2018). Furthermore, it is still unclear in that two subscales may benefit the clinical practices because some items of the two facets possessed strong correlation coefficients (Braithwaite et al., 2010). With these strong relationships, some items may not be distinguishable between two subscales such that it appears to undermine the construct validity of the tool. Hence, the given issue may also affect interpretation, monitoring, and management of the patients. Moreover, an obvious research gap arising from the START systematic review and meta-analysis revealed an insufficient application of the START in different contexts and populations, especially in Asian populations (Douglas, Pugh, Singh, Savulescu, & Fazel, 2017; O'Shea & Dickens, 2014).

Thus, the evidence on psychometric properties across cultural contexts, settings, and populations could support the generalizability of the tool. Overall, more research has been called for before established conclusion can be envisaged concerning the psychometric properties and generalizability of the START. In this regard, this research aimed to examined the psychometric properties of the START: Thai version, that is, the inter-rater reliability over mental health professionals across 118 Thai forensic psychiatric inpatients was investigated using the ICC and the construct validity of the Thai START in 118 Thai forensic psychiatric inpatients was assessed using Confirmatory Factor Analysis (CFA). However, the ordinal CFA with the Weighted Least Squares mean and Variance (WLSMV) adjusted estimator was used in the current study for dealing

with few categories of ordinal data from the Thai START scores (DiStefano & Morgan, 2014; Flora & Curran, 2004).

2. Materials and Methods

2.1 Participants and procedures

A total of 118 participants participated in this study and all participants were recruited from forensic psychiatric inpatients hospitalized in the forensic psychiatric unit of the Galya Rajanagarindra Institute (GYI) between January 2014 and January 2016. Inclusion and exclusion criteria were as follows: Eligible age for study was over 18 years and eligible gender was both male and female participants. The approval from the GYI Institutional Review Board (IRB No. 17/2557) was obtained before conducting the data collection.

Potential participants were obtained an information sheet to include i) a brief description of the study (i.e., objectives), ii) criteria used to determine eligibility, iii) a brief list of significant benefits and plausible risks (direct and/or indirect), iv) name and address of the principal researcher and workplace, and v) location of the research and name of the person to contact detail for further information. All participants signed the consent forms and agreed to participate in the study. In addition, all participants were ensured regarding the anonymity and confidentiality of their data.

The psychiatric clinical team, that is, seven senior psychiatrists, five clinical psychologists, eighty-five psychiatric nurses, two occupational therapists, and four social workers were voluntarily involved in the study and administrated the Thai START to each patient and then item scores for the Thai START across all the mental health professionals (maximum of five staff per a patient)

were derived by the median value. Before using the Thai START, all psychiatric clinical staff took part in the START training and workshop in the forensic psychiatric inpatient ward of the GYI, Bangkok. The training modules included violence risk assessment, the START-coding and practices, and also risk formulation of Thai forensic case studies.

2.2 Measure

Descriptive information on socio-demographic characteristics including age, gender, marital status and psychiatric characteristics, such as psychiatric diagnosis, number of previous hospitalizations and number of psychiatric emergency unit visits were collated from medical records.

The START consists of twenty dynamic items, that is, social skills, relationships, occupational, recreational, self-care, mental state, emotional state, substance use, impulse control, external triggers, social support, material resources, attitudes, medical adherence, rule adherence, conduct, insight, plans, coping, and treatability (Webster et al., 2009). Each item is rated for both degrees of strength and vulnerability on two three-point scales (0-1-2 format). Score 0 refers to no/minimal evidence of vulnerability or strength. Score 1 reflects moderate vulnerability or moderate strength. Score 2 represents high vulnerability or high strength (Webster et al., 2009).

The forward and backward translation procedure was adopted, and Thai senior psychiatrists and a professional native English-Thai bilingual and bicultural translator were involved. Some adaptation in wording was made in terms of functional equivalence of words (Hui & Triandis, 1985). The completed back-translation version was also rechecked by three authorities: two consultant forensic psychiatrists from Department of Forensic and Neurodevelopmental Science,

King's College London, and an original author of the START, Professor Johann Brink from University of British Columbia, Canada.

2.3 Statistical analyses

The IBM Statistical Package for the Social Sciences (SPSS) statistics 26 was used to analyze descriptive statistics (i.e., mean, median, standard deviation, and interquartile range). Inter-rater agreement among the psychiatric clinical team (i.e., psychiatrists, forensic nurses, clinical psychologists, social workers, and occupational therapists) was assessed using Intraclass Correlation Coefficient (ICC). To test the underlying structure or construct of the Thai START, the two-factor model with 20 indicators was examined. Strength and vulnerability subscales of the Thai START reflects the two-factor model and each subscale contains 20 test items (Gill Matthew & Brookes, 2017).

Mplus 7.0 software was employed to analyze the polychoric correlation coefficients among test items with ordinal data and to conduct CFA with robust WLSMV adjusted estimator (DiStefano & Morgan, 2014). The CFA model fit was evaluated by multiple fit indices, namely χ^2, χ^2/df, the Root Mean Square Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and Tucker Lewis index (TLI). The value of χ^2 greater than .05 shows good fit. The χ^2/df value less than 2.0 considers acceptable fit. The cut-off values for good model fits were adopted as RMSEA < .08, and CFI and TCI > .90 (Brown, 2015).
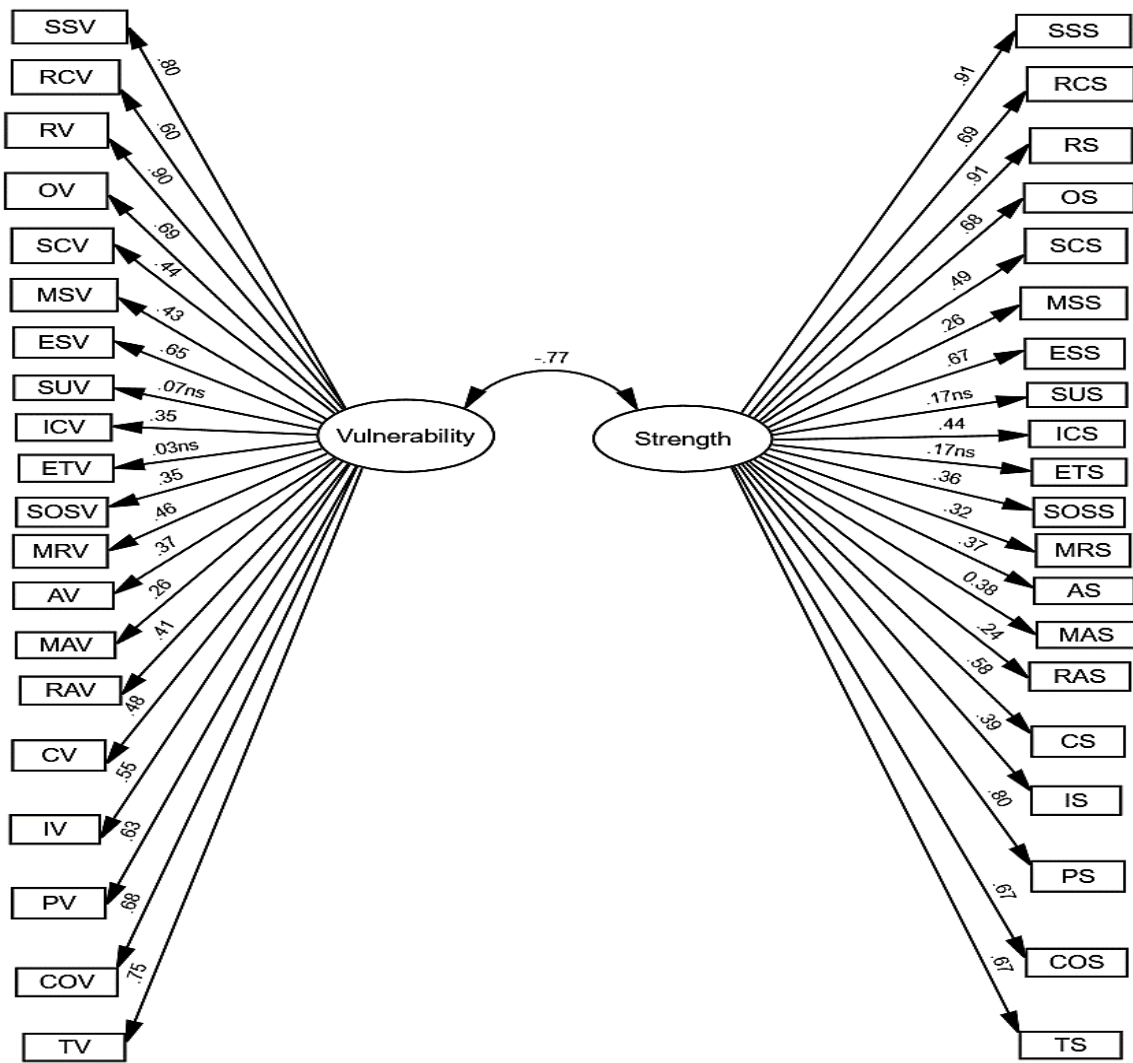
3. Results

Most of the patients were male (87.3%) and a large proportion of the participants were unemployed (44.1%), followed by employee (30.50) and merchant (8.5%), respectively. The commonest diagnoses were paranoid schizophrenia (37.2%), undifferentiated schizophrenia (17.8%) and adjustment disorder (9.4%), respectively. The mean age of participants was 39.4 years with standard deviation of 13.1 years. The youngest participant was 19 years old, and the oldest participant was 86 years old. The ICC values across test items showed a good to excellent reliability (.80 to .96) for the Thai START strength subscale and (.88 to .99) for the Thai START vulnerability subscale. The Thai START item scores were approximately normally distributed as indicated by skewness and kurtosis values (George & Mallery, 2010) (see **Table 1**.).

Table 1. ICC, Mean (SD), Median (IQR), Skewness, and Kurtosis across the Thai START items.

Thai START domains	ICC	Mean (SD)	Median (IQR)	Skewness	Kurtosis
Strength/Vulnerability					
Social skills (SSS/SSV)	.89/.91	1.07(.45)/1.14(.54)	1.00(.00)/1.00(.00)	.31/.11	1.95/.28
Relationships (RS/RV)	.80/.90	.98(.47)/.96(.56)	1.00(.00)/1.00(.00)	-.06/-.02	1.65/.25
Occupational (OS/OV)	.89/.94	.72(.63)/1.21(.74)	1.00(1.00)/1.00(1.00)	.28/-.36	-.63/-1.09
Recreational (RCS/RCV)	.88/.94	.81(.51)/.97(.64)	1.00(.00)/1.00(.00)	-.30/.03	.26/-.51
Self-care (SCS/SCV)	.96/.99	.97(.77)/1.14(.87)	1.00(2.00)/1.00(2.00)	.04/-.27	-1.29/-1.62
Mental state (MSS/MSV)	.94/.95	.68(.65)/1.42(.70)	1.00(1.00)/1.00(1.00)	.44/-.78	-.70/-.59
Emotional state (ESS/ESV)	.92/.94	.97(.61)/1.14(.68)	1.00(.00)/1.00(1.00)	.02/-.17	-.28/-.81
Substance use (SUS/SUV)	.93/.95	1.15(.65)/.97(.72)	1.00(1.00)/1.00(1.00)	-.16/.04	-.64/-1.06
Impulse control (ICS/ICV)	.93/.96	1.06(.60)/.99(.71)	1.00(.00)/1.00(1.00)	-.02/.01	-.21/-.99
External triggers (ETS/ETV)	.91/.93	.94(.59)/1.03(.65)	1.00(.00)/1.00(.00)	.01/-.03	-.07/-.61
Social support (SOSS/SOSV)	.95/.94	1.05(.65)/1.03(.65)	1.00(.00)/1.00(.00)	-.05/-.02	-.61/-.61
Material resources (MRS/MRV)	.94/.94	.84(.70)/1.19(.67)	1.00(1.00)/1.00(1.00)	.24/-.23	-.56/-.95
Attitudes (AS/AV)	.88/.95	.80(.52)/1.14(.77)	1.00(1.00)/1.00(1.00)	-.26/.53	-.75/.08
Medical adherence (MAS/MAV)	.85/.93	1.37(.49)/.75(.67)	1.00(1.00)/1.00(1.00)	.32/-.35	-1.25/-1.75
Rule adherence (RAS/RAV)	.92/.98	1.14(.77)/.52(.78)	1.00(1.00)/1.00(1.00)	1.09/-.36	-.76/-.85

Table 1. ICC, Mean (SD), Median (IQR), Skewness, and Kurtosis across the Thai START items.

Thai START domains	ICC	Mean (SD)	Median (IQR)	Skewness	Kurtosis
Conduct (CS/CV)	.93/.97	1.36(.61)/.61(.79)	1.00(1.00)/0.00(1.00)	.81/.36	-.47/-.65
Insight (IS/IV)	.92/.92	.66(.62)/1.36(.56)	1.00(1.00)/1.00(1.00)	-.14/.56	-.89/-.64
Plans (PS/PV)	.93/.92	.57(.61)/1.41(.63)	1.00(1.00)/1.00(1.00)	-.58/-.12	-.74/-.59
Coping (COS/COV)	.92/.89	.62(.52)/1.43(.50)	1.00(1.00)/1.00(1.00)	.28/.18	-1.18/-1.96
Treatability (TS/TV)	.89/.88	1.13(.52)/.95(.54)	1.00(.00)/1.00(.00)	.18/-.05	.55/.53

Figure 1. A path diagram for the two-factor model with 40 indicators and all factor loading coefficients were shown in standardized scores (ns = non-significant).

Figure 1 demonstrates the CFA result of the two-factor model representing the two Thai START subscales (i.e., strength vs vulnerability) of the Thai START with 40 indicators. The overall fit indices suggested an acceptable fit (N= 118; χ^2= 1205.15, *df*= 706, *p*<.01; χ^2/df = 1.71; RMSEA=.07; CFI=.90; TLI=.90). All factor loadings were statistically significant (*p*s<.05 and

.01) except SUV and ETV for the vulnerability subscale and SUS and ETS for the strength subscale. The relationship between strength and vulnerability factors was -.77 (*p*<.01).

4. Discussion

This study was conducted to examine the underlying structures of a translated version of the START within a Thai psychiatric institution. A Thai sample of 118 forensic psychiatric inpatients were enrolled in the study. It is evident in previous studies that risk assessment has been largely derived risk assessment instruments developed in Western clinical or criminal justice populations and also evaluated in similar settings (O'Shea & Dickens, 2014). This research investigated the generalizability of these advances in forensic mental health practice to an Asian country where cultural, criminal justice, and clinical factors differ markedly from the Western environment. The good-to-excellent inter-rater reliability was observed for the Thai START subscales and it also supports the previous claim of the strong reliability generalization of the tool (De Beuf, de Ruiter, Edens, & de Vogel, 2021). The current result rather supports the two-factor model of the START with a strong relationship between the strength and vulnerability factors. Although the Thai START demonstrated a good to excellent reliability and acceptable construct validity, substance use and external trigger items did not correlate with the subscales. This might have great clinical relevance for the selection of key items for predicting the violence behavior and treatment outcomes (Paetsch et al., 2019). It is also of note that this may uniquely reflect the perceptions of the Thai mental health professionals. An important contribution and implication of this study is that implementing the START in forensic setting should consider relevant cultural factors in a clinical environment where serious violence is rarer than in previous research for western samples (Zhou et al., 2016).

Further, previous research indicated duplicating and redundant constructs between the strength and vulnerability START subscales (Abidin et al., 2013; O'Shea & Dickens, 2014; Wilson et al., 2010). In contrast, this study is in line with several studies that both subscales are distinguishable, and they should be retained because of the potential to inform therapeutic relationships and risk management plans (De Beuf et al., 2020; de Ruiter & Nicholls, 2011; Nonstad et al., 2010). Thus, this finding suggests a unique contribution of each subscale to the Thai START. It is also plausible that specific variation in each subscale may have differential predictive power over time, and this is a matter for possible future investigation.

To our knowledge, this is the first study that has examined the factorial structure of the START using CFA. Nonetheless, the study presents some limitation. Although our sample insisted of 118 participants with neurological and psychiatric disorders (e. g. schizophrenia and adjustment disorder), the current results cannot be generalized to other forensic psychiatric samples. It would be beneficial to test the invariance of the scale across gender. In the current study, the small subsample size did not allow us to test measurement invariance with a sufficient statistical power. Finally, further research with a larger sample is required to improve a generalizability of the START.

Author Contributions: Conceptualization: WS. Data curation: WS. Formal analysis: WS, PW. Investigation: WS. Methodology: WS. Project administration: WS. Resources: WS. Software: WS. Supervision: WS. Validation: WS. Writing-original draft: WS. Writing-review & editing: WS, PW.

16

Institutional Review Board Statement: This study design was approved by the GYI Institutional Review Board (IRB No. 17/2557).

Informed Consent Statement: The informed consent was obtained from all participants participated in the current study.

Conflicts of Interests: The authors have no potential conflicts of interest to disclose.

References

Abidin, Z., Davoren, M., Naughton, L., Gibbons, O., Nulty, A., & Kennedy, H. G. (2013). Susceptibility (risk and protective) factors for in-patient violence and self-harm: prospective study of structured professional judgement instruments START and SAPROF, DUNDRUM-3 and DUNDRUM-4 in forensic mental health services. *BMC Psychiatry, 13*(1), 197. doi:10.1186/1471-244X-13-197

Atkinson, T. M., Rosenfeld, B. D., Sit, L., Mendoza, T. R., Fruscione, M., Lavene, D., . . . Basch, E. (2011). Using confirmatory factor analysis to evaluate construct validity of the Brief Pain Inventory (BPI). *Journal of pain and symptom management, 41*(3), 558-565. doi:10.1016/j.jpainsymman.2010.05.008

Braithwaite, E., Charette, Y., Crocker, A. G., & Reyes, A. (2010). The Predictive Validity of Clinical Ratings of the Short-Term Assessment of Risk and Treatability (START). *International Journal of Forensic Mental Health, 9*(4), 271-281. doi:10.1080/14999013.2010.534378

Brown, T. A. (2015). *Confirmatory factor analysis for applied research, 2nd ed*. New York, NY, US: The Guilford Press.

Cartwright, J. K., Desmarais, S. L., Hazel, J., Griffith, T., & Azizian, A. (2018). Predictive validity of HCR-20, START, and static-99R assessments in predicting institutional aggression among sexual offenders. *Law and Human Behavior, 42*(1), 13-25. doi:10.1037/lhb0000263

De Beuf, T. L. F., de Ruiter, C., & de Vogel, V. (2020). Staff Perceptions on the Implementation of Structured Risk Assessment with the START:AV: Identifying Barriers and Facilitators in a Residential Youth Care Setting. *International Journal of Forensic Mental Health*, 1-18. doi:10.1080/14999013.2020.1756994

De Beuf, T. L. F., de Ruiter, C., Edens, J. F., & de Vogel, V. (2021). Taking "the boss" into the real world: Field interrater reliability of the Short-Term Assessment of Risk and Treatability: Adolescent Version. *Behavioral Sciences & the Law, n/a*(n/a). doi:https://doi.org/10.1002/bsl.2503

de Ruiter, C., & Nicholls, T. L. (2011). Protective Factors in Forensic Mental Health: A New Frontier. *International Journal of Forensic Mental Health, 10*(3), 160-170. doi:10.1080/14999013.2011.600602

Desmarais, S. L., Nicholls, T. L., Read, J. D., & Brink, J. (2010). Confidence and accuracy in assessments of short-term risks presented by forensic psychiatric patients. *The Journal of Forensic Psychiatry & Psychology, 21*(1), 1-22. doi:10.1080/14789940903183932

Desmarais, S. L., Nicholls, T. L., Wilson, C. M., & Brink, J. (2012). Using dynamic risk and protective factors to predict inpatient aggression: reliability and validity of START assessments. *Psychological Assessment, 24*(3), 685-700. doi:10.1037/a0026668

DiStefano, C., & Morgan, G. B. (2014). A Comparison of Diagonal Weighted Least Squares Robust Estimation Techniques for Ordinal Data. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(3), 425-438. doi:10.1080/10705511.2014.915373

Douglas, T., Pugh, J., Singh, I., Savulescu, J., & Fazel, S. (2017). Risk assessment tools in criminal justice and forensic psychiatry: The need for better data. *European psychiatry : the journal of the Association of European Psychiatrists, 42*, 134-137. doi:10.1016/j.eurpsy.2016.12.009

Flora, D. B., & Curran, P. J. (2004). An Empirical Evaluation of Alternative Methods of Estimation for Confirmatory Factor Analysis With Ordinal Data. *Psychological Methods, 9*(4), 466-491. doi:10.1037/1082-989X.9.4.466

George, D., & Mallery, P. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference 17.0 Update* (10 ed.). Boston: Pearson.

Gill Matthew, J., & Brookes, S. (2017). The psychology quadrant: an outcome measure based on the START risk assessment. *Journal of Forensic Practice, 19*(1), 14-22. doi:10.1108/JFP-03-2016-0018

Haines, A., Brown, A., Javaid, S. F., Khan, F., Noblett, S., Omodunbi, O., . . . Whittington, R. (2018). Assessing Protective Factors for Violence Risk in U.K. General Mental Health Services Using the Structured Assessment of Protective Factors. *International Journal of Offender Therapy and Comparative Criminology, 62*(12), 3965-3983. doi:10.1177/0306624x17749449

Hogan, N. R., & Olver, M. E. (2018). A Prospective Examination of the Predictive Validity of Five Structured Instruments for Inpatient Violence in a Secure Forensic Hospital. *International Journal of Forensic Mental Health, 17*(2), 122-132. doi:10.1080/14999013.2018.1431339

Hui, C. H., & Triandis, H. C. (1985). Measurement in Cross-Cultural Psychology:A Review and Comparison of Strategies. *Journal of Cross-Cultural Psychology, 16*(2), 131-152. doi:10.1177/0022002185016002001

Lowder, E. M., Desmarais, S. L., Rade, C. B., Johnson, K. L., & Van Dorn, R. A. (2019). Reliability and Validity of START and LSI-R Assessments in Mental Health Jail Diversion Clients. *Assessment, 26*(7), 1347-1361. doi:10.1177/1073191117704505

McArdle, J. J. (1996). Current Directions in Structural Factor Analysis. *Current Directions in Psychological Science, 5*(1), 11-18. doi:10.1111/1467-8721.ep10772681

Messick, S. (1989). Validity. In R. Linn (Ed.), *Educational measurement* (pp. 13-104). New York: American Council on Education & Macmillan Publishing.

Nicholls, T. L., Brink, J., Desmarais, S. L., Webster, C. D., & Martin, M.-L. (2006). The Short-Term Assessment of Risk and Treatability (START):A Prospective Validation Study in a Forensic Psychiatric Sample. *Assessment, 13*(3), 313-327. doi:10.1177/1073191106290559

Nonstad, K., Nesset, M. B., Kroppan, E., Pedersen, T. W., Nøttestad, J. A., Almvik, R., & Palmstierna, T. (2010). Predictive Validity and Other Psychometric Properties of the Short-Term Assessment of Risk and Treatability (START) in a Norwegian High Secure Hospital. *International Journal of Forensic Mental Health, 9*(4), 294-299. doi:10.1080/14999013.2010.534958

O'Shea, L. E., & Dickens, G. L. (2014). Short-Term Assessment of Risk and Treatability (START): Systematic review and meta-analysis. *Psychological Assessment, 26*(3), 990-1002. doi:10.1037/a0036794

O'Shea, L. E., Picchioni, M. M., & Dickens, G. L. (2016). The Predictive Validity of the Short-Term Assessment of Risk and Treatability (START) for Multiple Adverse Outcomes in a Secure Psychiatric Inpatient Setting. *Assessment, 23*(2), 150-162. doi:10.1177/1073191115573301

Paetsch, A., van Os, T. W. D. P., Troquete, N. A. C., & van den Brink, R. H. S. (2019). Single-item predictive validity of the Short-Term Assessment of Risk and Treatability (START) for violent behaviour in outpatient forensic psychiatry. *The Journal of Forensic Psychiatry & Psychology, 30*(4), 630-641. doi:10.1080/14789949.2019.1581831

Rogers, R. (2000). The Uncritical Acceptance of Risk Assessment in Forensic Practice. *Law and Human Behavior, 24*(5), 595-605. doi:10.1023/A:1005575113507

Singh, J. P., Desmarais, S. L., Hurducas, C., Arbach-Lucioni, K., Condemarin, C., Dean, K., . . . Otto, R. K. (2014). International Perspectives on the Practical Application of Violence Risk Assessment: A Global Survey of 44 Countries. *International Journal of Forensic Mental Health, 13*(3), 193-206. doi:10.1080/14999013.2014.922141

Strauss, M. E., & Smith, G. T. (2009). Construct validity: advances in theory and methodology. *Annual review of clinical psychology, 5*, 1-25. doi:10.1146/annurev.clinpsy.032408.153639

Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *Int J Med Educ, 11*, 245-247. doi:10.5116/ijme.5f96.0f4a

Viljoen, J. L., Beneteau, J. L., Gulbransen, E., Brodersen, E., Desmarais, S. L., Nicholls, T. L., & Cruise, K. R. (2012). Assessment of Multiple Risk Outcomes, Strengths, and Change with the START:AV: A Short-Term Prospective Study with Adolescent Offenders. *International Journal of Forensic Mental Health, 11*(3), 165-180. doi:10.1080/14999013.2012.737407

Viljoen, S., Nicholls, T., Greaves, C., de Ruiter, C., & Brink, J. (2011). Resilience and Successful Community Reintegration among Female Forensic Psychiatric Patients: A Preliminary Investigation. *Behavioral Sciences & the Law, 29*(5), 752-770. doi:https://doi.org/10.1002/bsl.1001

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Desmarais, S. L. (2009). *Manual for the Short Term Assessment of Risk and Treatability (START)(Version 1.)*. BC Mental Health, St. Joseph's Healthcare, British Columbia.

Webster, C. D., Martin, M. L., Brink, J., Nicholls, T. L., & Middleton, C. (2004). *Manual for the Short Term Assessment of Risk and Treatability (START). Version 1.0, Consultation edition*. St. Joseph's Healthcare, Hamilton, Ontario, Canada, and Forensic Psychiatric Services Commission, Port Coquitlam, British Columbia, Canada.

Whittington, R., Bjørngaard, J. H., Brown, A., Nathan, R., Noblett, S., & Quinn, B. (2014). Dynamic relationship between multiple START assessments and violent incidents over time: a prospective cohort study. *BMC Psychiatry, 14*(1), 323. doi:10.1186/s12888-014-0323-7

Wilson, C. M., Desmarais, S. L., Nicholls, T. L., & Brink, J. (2010). The Role of Client Strengths in Assessments of Violence Risk Using the Short-Term Assessment of Risk and

Treatability (START). *International Journal of Forensic Mental Health, 9*(4), 282-293. doi:10.1080/14999013.2010.534694

Zhou, J., Witt, K., Xiang, Y., Zhu, X., Wang, X., & Fazel, S. (2016). Violence risk assessment in psychiatric patients in China: A systematic review. *Aust N Z J Psychiatry, 50*(1), 33-45. doi:10.1177/0004867415585580