Research Brief
Internal Structure of the Static Risk Offender Need Guide for Recidivism (STRONG-R)

Xiaohan Mei, M.A.
Zachary Hamilton, Ph.D.
Washington State University

STRONG-R and Internal Structure

Construct validity is the degree of accuracy an instrument measures the targeted constructs. To claim construct validity, an assessment instrument must demonstrate subsequent criterion-related validities, such as convergent and discriminant/divergent validity, predictive validity, an evidence-based factor structure, as well as content validity and reliability (Messick, 1993; Hayes, Richard & Kubany, 1995). Beside these traditional types of evidence, an assessment instrument's internal structure is considered key evidence of construct validity (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). The current research examines the internal latent factor structure of the Static Risk Offender Need Guide – Revised (STRONG-R) instrument. To establish an instrument's latent structure three basic aspects must be identified: dimensionality, measurement invariance, and reliability (Rios and Wells, 2014).

Dimensionality

The common method for establishing dimensional validation is Confirmatory Factor Analysis (CFA) for first order latent construct/sub-scales. For constructs with many dimensions, the common approaches are correlated-factors (measurement model) for second-order models, as well as bi-factor models within the framework of Structural Equation Modeling (SEM) (Rios & Wells, 2014). Specifically, model fit indices and (standardized) factor loadings are common criterions when examining the dimensionality aspects of the internal structure (Bollen, 1989; Brown, 2014; Kline, 2010; Thompson, 2004; Little, 2013). CFA is used to confirm the hypothesized number of factors based on model fit indices. If multiple factors are identified and factors are substantially correlated with one and another, a higher order factor is used to explain the common variance among the lower factor.

Measurement Invariance

From a psychometric perspective, measurement invariance represents a lack of systematic bias resulting from group membership, which may lead to a misinterpretation of scores. Measurement invariance is achieved at both the item and scale level. Item invariance is achieved when all measures within a dimension have the same item functioning. Items that do not achieve invariance are often referred to as differential item functioning (or DIF) (Sireci & Rios, 2013). Scale-level invariance concerns the invariance level across different groups. Such hypothesized invariance is analyzed by using Multiple Group Confirmatory Factor Analysis (MGCFA) (Rios & Wells, 2014). The benefits of MGCFA include: (1) model fit statistics simultaneously evaluated across multiple groups; (2) testing model invariant levels; and (3) between groups true score mean difference estimates. MGCFA identifies if a model possesses one of three levels of invariance: configural (measurement model), metric (tau-equivalent across groups), and scale (equivalent intercept or thresholds) (Dimitrov, 2010). If a higher order factor/construct is extracted, then additional tests are used to examine the higher-order measurement invariances.

Method

The current research was conducted within the framework of Item Response Theory (IRT) and Item Factor Analysis (IFA). IRT and IFA are developed in relative isolation, but the difference in their forms is, at best, superficial (Thomas, 2011). IRT and/or IFA was used as the

majority of the items in STRONG-R are categorical/ordinal in nature. It should be noted, gender was used as group membership when conducting MGCFA, meaning the risk and needs of both male and female populations could be fairly estimated without bias of group membership.

The current study was conducted using a series of psychometric tests for dimensionality and measurement invariance. First, MGCFA at the configural level was conducted, then scale invariant robust tests were completed, and tau-equivalent constraints were imposed in IFA models. When factors correlate substantially, higher level models were constructed, and measurements invariance tests were conducted in second order models. All levels of factor-level loadings were constrained to be equal from top-to-bottom within and across groups, and all manifest items were constrained to be equal within-and-across groups[1].

Once the internal structure is confirmed, scale invariance and item invariance are achieved. An IRT model becomes a Rasch Model and is often referred to as one-parameter model, which simplifies the calculation (Thomas, 2011). In Rasch Models, total scores provide sufficient statistical properties to establish individuals' latent trait, and Cronbach's Alpha is considered an appropriate statistical test of scale reliability (Thomas, 2011)[2].

Finally, the composite scores of the five constructs were computed and used as indicators of offenders' global risk-needs estimates. These five indicators/items were used for the CFA model to examine the risk and needs construct/scales[3]. We use the common/true variance instead of unique variance within each scale in the model in order to confirm the conceptualized and hypothesized internal and latent structural pattern[4]. This is achieved by multiplying the total variance of the construct with its Cronbach's alpha coefficients before conducting analysis in the final model[5].

Results

As showed in Tables 1, 2, 3, 4 and 5. All five constructs, independently, passed group invariance, higher order modeling, and scale invariance and item invariance tests. Specifically, when imposing constrains, the CFI and TLI value change met the industry standard of less than .01. Based on these findings, several conclusions could be drawn. First, all manifest items represented fair tests for both male and female offenders, as all items demonstrated statistically equivalent discrimination and threshold values. Next, with second order modeling, the internal latent structure for each construct was confirmed. Third, total scores of first order factor/sub-scale provided legitimate estimate of offenders' risks and needs.

---

[1] When examining the quality of the items and factor models, we utilized industry accepted standards and model fit indices, including the Comparative Fit Indexes (CFI), the Tucker-Lewis Index (TLI) and the Root Mean Square Error of Approximation (RMSEA) model fit indices. Evaluating measurement invariance requires imposing sequential constraint on model parameters, and the levels of measurement invariance comparisons were conducted via the ΔCFI and ΔTLI index (Little, 2014).

[2] When estimating assessment's social-psychometric properties, IRT-Rasch Models demand all items possess the same discrimination values (or IFA loading). In the IRT/IFA framework, both latent traits (such as anti-social cognition) and those assessed characteristics of the items are assumed to dictate a person's response to assessed items (Thomas, 2011). The total scores of the scales of the five constructs were then obtained, and the convergent and divergent validity of the STRONG-R will be examined via conducting an Exploratory Factor Analysis (Mei, Routh, & Hamilton, 2016a).

[3] This is also referred to as the data reduction technique as described by Brown (2012).

[4] It should be noted, this test was conducted within the Classical Test Theory (CTT) framework, and total scores were only sufficient estimates for first order and/or Rash Models but not for higher order models. The total score of a second/higher order construct not only contains common and true score variance among the first order/lower order factors, but also factor disturbance, which is the unique variance not accounted for by the higher order factors.

[5] In psychometric analysis, this is also referred to as creating "phantom variables".

Table 1 Measurement Invariance and Structural Tests – Anti-Social History

Model	Test of Structure/Invariance	*df*	CFI	TLI	RMSEA [90% C.I.]	Δ *df*	Δ CFI	Δ TLI	Pass?
A1	Measurement Model - Configural Invariant	142	.936	.917	.070 [.070 - .071]	--	--	--	--
A2	Measurement Model - Strong Invariant	171	.940	.936	.062 [.061 - .063]	29	+.004	+.019	Yes
A3	First Order Factor Item Invariant + Group Invariant	181	.939	.939	.061 [.060 - .061]	10	.001	+.003	Yes
A4*	Second Order Baseline Model	185	.934	.935	.063 [.062 - .063]	--	--	--	--
A5	Second Order Tau Equivalent + Gender Invariant	191	.935	.938	.061 [.060 - .062]	6	+.001	+.003	Yes
A6*	Third Order Baseline Model	193	.931	.935	.063 [.062 - .063]	--	--	--	--
A7*	Third Order Tau Equivalent + Gender Invariant	194	.933	.937	.061 [.061 - .062]	1	+.002	+.002	Yes
A8 to A6	Third Order Tau Equivalent + Gender Invariant	193	.935	.939	.061 [.060 - .061]	0	+.002	+.002	Yes
A9	All First Order Item Invariant + Gender Invariant	197	.930	.936	.062 [.061 - .063]	4	.005	.003	Yes
A10	All Second Order Tau Equivalent + Gender Invariant	198	.920	.927	.066 [.066 - .067]	1	.010	.009	Yes

*Factor mean not estimated

As presented in Table 1, in Anti-social History Scale, a third order factor model was confirmed. Group/Gender invariance was achieved. The model passed tau-equivalent tests at item level, second order and third order level. Anti-social History, as operationalized, is the extent of engagement in prior anti-social behavior. It contains two scales, including Criminal History and Violence History, measuring individuals' Adult Criminal History, Prison Infraction History, Domestic and Misdemeanor Violence History, and Aggression History (visual representation of the construct, see Figure 1).

Figure 1. Anti-social History

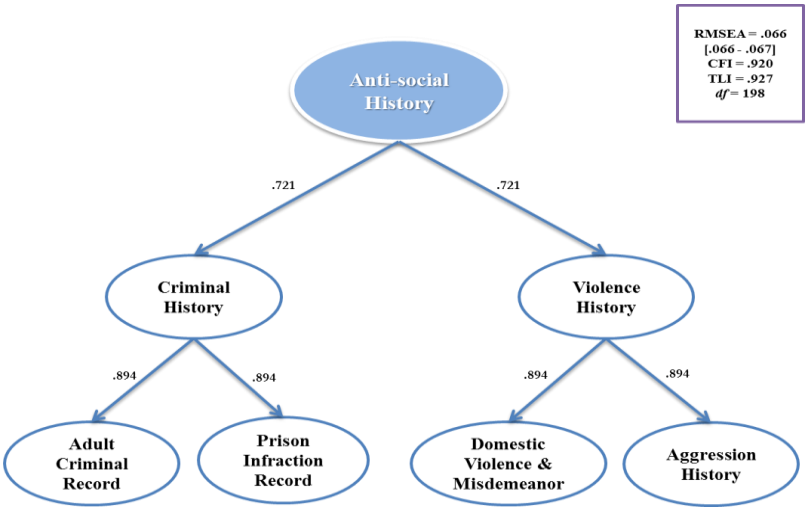


Table 2 Measurement Invariance and Structural Tests – Education & Employment

Model	Test of Structure/Invariance	*df*	CFI	TLI	RMSEA [90% C.I.]	Δ *df*	Δ CFI	Δ TLI	Pass?
B1	Measurement Model - Configural Invariant	88	.988	.982	.037 [.036 - .039]	--	--	--	--
B2	Measurement Model - Strong Invariant	115	.983	.981	.038 [.037 - .039]	27	.005	.001	Yes
B3	First Order Factor Item Invariant + Group Invariant	122	.980	.978	.041 [.040 - .042]	7	.003	.003	Yes
B4*	Second Order Baseline Model	127	.970	.969	.049 [.048 - .050]	--	--	--	--
B5	Second Order Tau Equivalent + Gender Invariant	141	.964	.966	.051 [.050 - .052]	14	.006	.003	Yes
B6*	Third Order Baseline Model	143	.962	.965	.052 [.051 - .053]	--	--	--	--
B7*	Third Order Tau Equivalent + Gender Invariant	144	.962	.965	.052 [.051 - .053]	1	.000	.000	Yes
B8 to B6	Third Order Tau Equivalent + Gender Invariant	143	.960	.963	.053 [.053 - .054]	0	.002	.002	Yes
B9	Full First + Second Order Tau Equivalent + Gender Invariant	147	.955	.959	.056 [.055 - .057]	4	.005	.004	Yes

*Factor mean not estimated

As displayed in Table 2, in Education/Employment Scale, a third order factor model was confirmed. The model passed tau-equivalent tests at item level, second order level and third order level. It also passed the gender invariance tests at all levels. Education/Employment is the degree of which an offender's susceptibility to engage in anti-social/criminal behaviors. It contains two scales, individuals' Education and Employment, and the legitimacy of their Income. Education and Employment Experience measures individuals achieved education level, legal working experience, and the extent of juvenile anti-social record as a diversion from the normal education and working trajectory. Income source measures the extent to which how much income an individual receives from either legal employment or illegal criminal actives. A visual representation of the construct is displayed in Figure 2).

Figure 2. Employment & Education

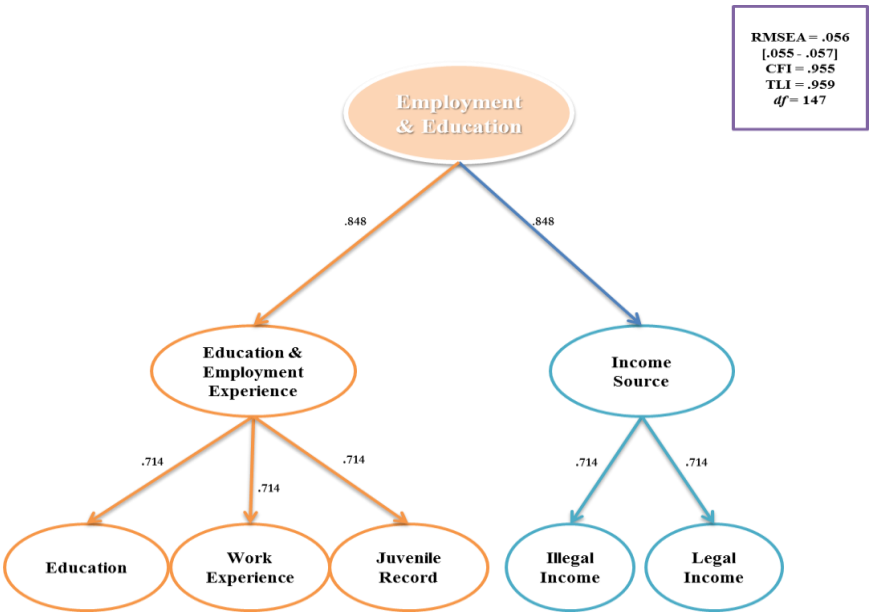


Table 3. Measurement Invariance and Structural Tests –Anti-Social Propensity

Model	Test of Structure/Invariance	*df*	CFI	TLI	RMSEA [90% C.I.]	Δ *df*	Δ CFI	Δ TLI	Pass?
C1	Measurement Model - Configural Invariant	1666	.978	.974	.018 [.018 - .018]	--	--	--	--
C2	Measurement Model - Strong Invariant	1694	.978	.974	.018 [.018 - .018]	28	.000	.000	Yes
C3	First Order Factor Item Invariant + Gender Invariant	1723	.972	.968	.020 [.020 - .020]	29	.006	.006	Yes
C4*	Second Order Baseline Model	1739	.969	.964	.021 [.021 - .021]	--	--	--	--
C5	Second Order Model - 4 Second Order Factors	1931	.962	.961	.022 [.022 - .023]	192	.007	.003	Yes
C6	Second Order Model Tau Equivalent + Gender Invariant	1959	.958	.958	.023 [.023 - .024]	28	.004	.003	Yes
C7*	Third Order Baseline Model	1962	.956	.956	.024 [.024 - .024]	--	--	--	--
C8	Third Order Model Tau Equivalent + Gender Invariant	1972	.952	.952	.025 [.025 - .025]	10	.004	.004	Yes
C9	All First Order Item Invariant + Gender Invariant	1987	.946	.947	.026 [.026 - .026]	15	.006	.005	Yes
C10	Second Order Tau Equivalent Across 2 Factors + Gender Invariant	1988	.941	.941	.027 [.027 - .028]	1	.005	.006	Yes
C11	Second Order Tau Equivalent Across 3 Factors + Gender Invariant	1989	.935	.935	.029 [.029 - .029]	1	.006	.006	Yes
C12	All (4) Second Order Tau Equivalent + Gender Invariant	1990	.934	.934	.029 [.029 - .029]	1	.001	.001	Yes

*Factor mean not estimated

As showed in Table 3, the Anti-Social Propensity construct confirmed as a third order factor model, achieved gender invariance, and passed all invariance tests at all levels. Anti-social Propensity assesses the extent to which an offender is likely to engage in anti-social behaviors. Anti-social Propensity contains four scales, including Anti-social Influence, Anti-social Cognition, Anti-social Personality and Violence Propensity. Anti-social Influence (contains three sub-scales) measures the degree to which an individuals' pro-criminal attitude, belief and values are the consequences of their association with their anti-social friends, family member and partner; anti-social cognition (contains four sub-scales) is the extent to which individuals' possess anti-social attitudes, beliefs, and values that are favorable to crime; anti-social personality (contains four sub-scales) is the extent to which individuals' demonstrated anti-social behavioral pattern; violence propensity (contains five sub-scales) assesses an offender's aggression issues and relative levels of aggressive behaviors/attitudes. A visual representation of the construct is provided in Figure 3).

Figure 3. Anti-social Propensity

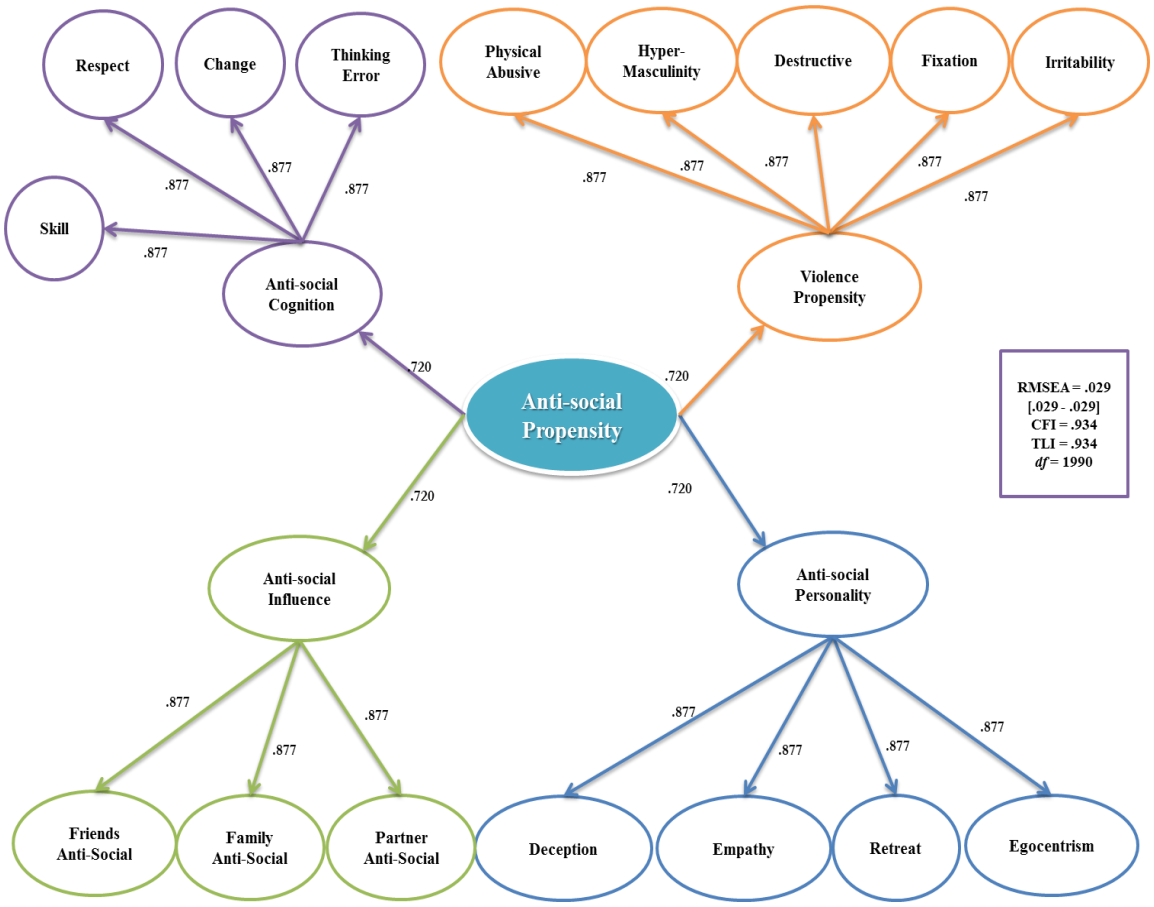


Table 4 Measurement Invariance and Structural Tests – Substance Abuse

Model	Test of Structure/Invariance	*df*	CFI	TLI	RMSEA [90% C.I.]	Δ *df*	Δ CFI	Δ TLI	Pass?
D1	Measurement Model - Configural Invariant	48	.948	.937	.066 [.064 - .067]	--	--	--	--
D2	Measurement Model - Strong Invariant	58	.959	.949	.059 [.057 - .060]	10	+.011	+.012	Yes
D3	First Order Factor Item Invariant + Gender Invariant	64	.954	.948	.060 [.058 - .061]	6	.005	.004	Yes
D4*	Second Order Baseline Model	67	.941	.937	.066 [.064 - .067]	--	--	--	--
D5	Second Order Model Tau Equivalent + Gender Invariant	71	.953	.952	.057 [.056 - .058]	4	+.012	+.015	Yes
D6	All First Order Item Invariant + Gender Invariant	73	.944	.945	.061 [.060 - .063]	2	.009	.007	Yes

*Factor mean not estimated

As presented in Table 4, the Substance Abuse construct was confirmed as a second order factor model, which achieved gender invariance and passed the tau-equivalent tests at all three levels. The Substance Abuse construct is comprised of three scales, including Hard Drug Use, Drug Use-Share-Barter, and Drug Related Crime. Hard Drug Use assess the extent to which an individual uses hard drugs; Drug Use-Share-Barter assesses the extent to which an offender's substance use is collective, or used with associates; Drug Related Crime assesses to what extent the offender is involved in drug-related crimes. A visual representation of the construct is provided in Figure 4).

Figure 4. Substance Abuse

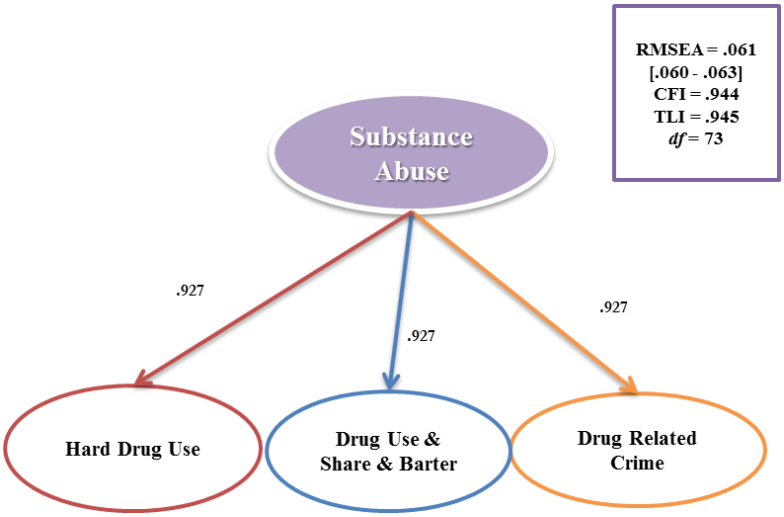


Table 5 Measurement Invariance and Structural Tests – Reintegration Needs

Model	Test of Structure/Invariance	*df*	CFI	TLI	RMSEA [90% C.I.]	Δ *df*	Δ CFI	Δ TLI	Pass?
E1	Measurement Model - Configural Invariant	556	.965	.959	.038 [.037 - .038]	--	--	--	--
E2	Measurement Model - Strong Invariant	576	.966	.962	.036 [.036 - .037]	20	+.001	+.003	Yes
E3	First Order Factor Item Invariant + Gender Invariant	595	.962	.959	.038 [.037 - .038]	19	.004	.003	Yes
E4*	Second Order Baseline Model	602	.962	.959	.038 [.037 - .038]	--	--	--	--
E5	Second Order Model Tau Equivalent + Gender Invariant	634	.954	.953	.040 [.040 - .041]	32	.008	.006	Yes
E6*	Third Order Baseline Model	636	.954	.953	.040 [.041 - .041]	2	.000	.000	Yes
E7	Second Order Model Tau Equivalent + Gender Invariant	640	.956	.956	.039 [.039 - .040]	4	+.002	+.003	Yes
E8	Full First + Second Order Tau Equivalent + Gender Invariant	647	.953	.953	.040 [.040 - .041]	7	.003	.003	Yes

*Factor mean not estimated

As displayed in Table 5, the Reintegration Needs construct was confirmed as third order model. It passed gender invariance tests and tau-equivalent tests at first order/item, second order and third order level. Reintegration Needs subsumes three scales, namely, Employment Barrier, Reentry Needs, and Mental Health. Employment Barrier measures the extent to which the obstacles (including, Mental/Physical, Motivation/Skills, Work Ethics and Systematic Barriers) individuals have to overcome in order to be successful in the reintegration process; Reentry Needs assesses assess the degree to which an offender needs additional help from social and justice system to meet daily, basic needs; Mental Health, which contains two sub-scales is the degree to which an offender's mental health condition (visual representation of the construct, see Figure 5).

Figure 5. Reintegration Needs

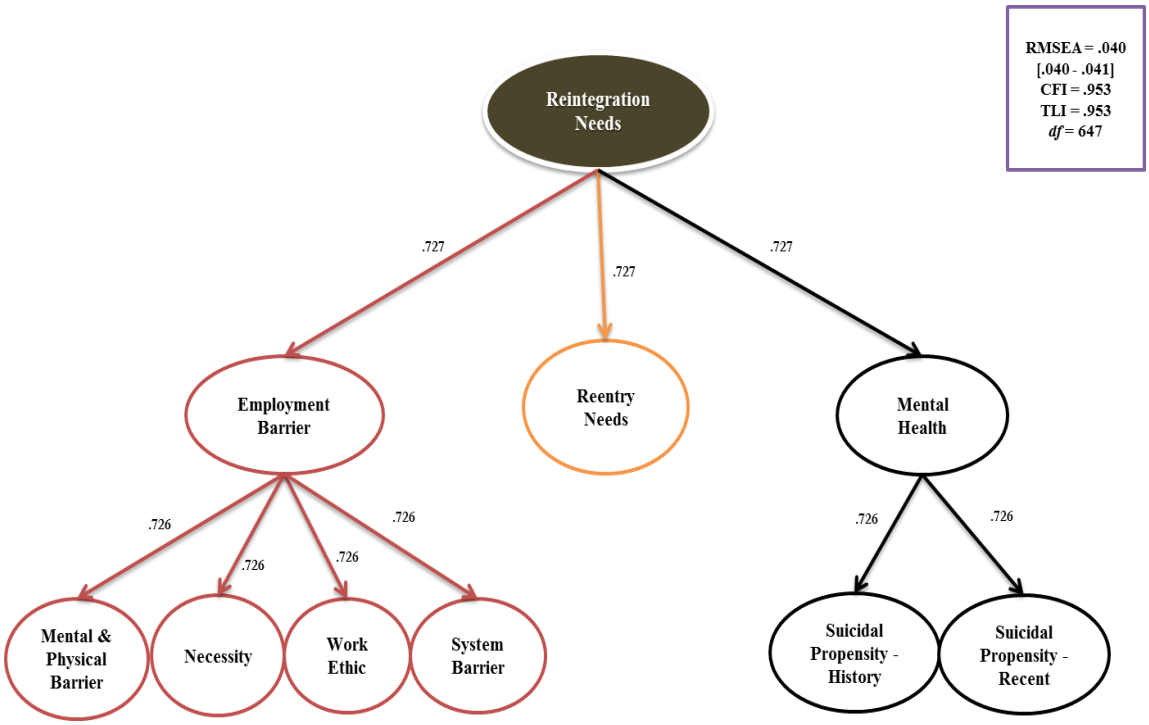
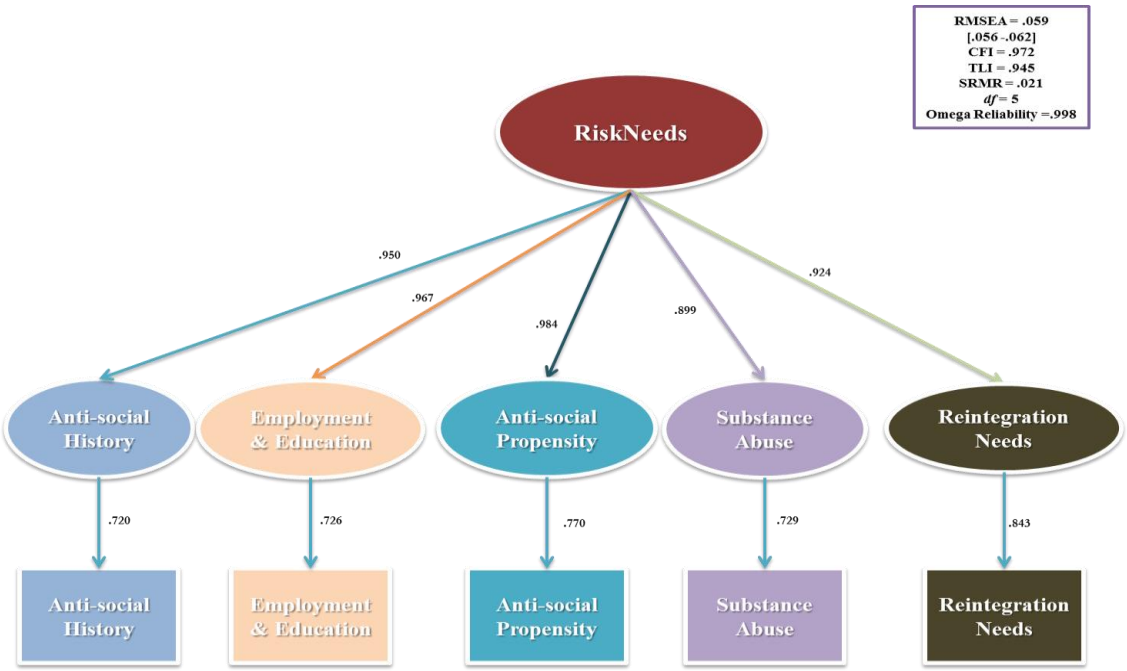


Table 6 Final Global Risk-Needs Model

Construct	No. of Item	Loading of Unweighted Total Score	Loading in the Final Model	Cronbach's Alpha	Model Fit Indicies
Anti-Social History	14	.445	.950	.720	**CFI = .972**
Education & Employment	8	.549	.967	.723	**TLI = .945**
Anti-Social Propensity	46	.643	.984	.770	***df* = 5**
Substance Abuse	9	.543	.899	.729	**RMSEA = .059 [.056 - .062]**
Reintegration Needs	26	.596	.924	.843	**SRMR = .021**
Global Risk-Needs	103	**Omega Reliability = .682**	**Omega Reliability = .998**	**Cronbach's Alpha = .658**	

Table 6 displays the results of the global risk-needs CFA using the unweighted total scores of the five constructs and variables with only true score variance that represent the five constructs. All model fit indices, including CFI, TLI, RMSEA and SRMR, indicate a good model fit. Loadings ranged from .445 to .643, which are relative weak but expected as the total scores of the five construct indicators contain common/true variance and the lower-order factors' unique variance/disturbance. However, when indicators' true score variance is used, the result of the model fit did not change, and the loadings range from .899 to .984. This finding confirmed the theorized internal structure of the STRONG-R.

Figure 6. Global Risk-Needs Model



Conclusion

In summary, the findings of psychometric tests of measurement invariance and higher order modeling provides sufficient evidence to support the theorized STRONG-R's internal latent structure, in which five higher order constructs were confirmed. As shown the STRONG-R internal latent factorial pattern displays a variety of individuals' risk and needs factors. To date, STRONG-R's reliability, content validity, convergent/discriminate validity, and concurrent/predictive validity have also been rigorously examined (Mei & Hamilton, 2016a; Mei & Hamilton, 2016b; Mei, Routh & Hamilton, 2016a; Mei, Routh & Hamilton, 2016b). Although, additional tests are needed going forward, we believe that we have amassed a considerable amount of empirical evidence to claim the construct validity of the STRONG-R.

REFERENCES

American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Bollen, K.A. (1989). *Structural equation models with latent variables*. New York, NY: Wiley.

Brown, T.A. (2014). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.

Dimitrov, D. M. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121-149.

Geiser, C. (2013). *Data analysis with Mplus*. Guildford: New York.

Haynes, S., Richard, D., Kubany, E., & Butcher, James N. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment*, *7*(3), 238-247.

Kline, R.B. (2010). *Principles and practice of structural equation modeling* (3rd ed.). New York, NY: Guilford Press. Little, T. D. (2013). *Longitudinal Structural Equation Modeling*. Guilford: New York.

Mei, X. & Hamilton, Z. (2016a). Reliability of the STRONG-R. Spokane, WA: Washington State Institute for Criminal Justice.

Mei, X. & Hamilton, Z. (2016b). Concurrent Validity of the STRONG-R. Spokane, WA: Washington State Institute for Criminal Justice.

Mei, X. Routh, D. & Hamilton, Z. (2016a). Convergent and Discriminate Validity of the STRONG-R. Spokane, WA: Washington State Institute for Criminal Justice.

Mei, X. Routh, D. & Hamilton, Z. (2016b). Content Validity of the STRONG-R. Spokane, WA: Washington State Institute for Criminal Justice.

Messick, S. (1993). Foundations of validity: Meaning and consequences in psychological assessment. *Ets research report series, 1993*(2), 1-18.

Rios, J, & Wells, C. (2014). Validity evidence based on internal structure. *Psicothema*, *26*(1), 108-116.

Sireci, S, & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, *26*(1), 100-107.

Thompson, M. (2004). The value of item response theory in clinical assessment: A review. *Assessment*, *18* (3), 291-307.

For further details about the STRONG-R internal structure research findings, WSU Researchers can be contacted at zachary.hamilton@wsu.edu

Suggested Citation: Mei, X. & Hamilton, Z. (2016). *Internal Structure of the STRONG-R*. Spokane, WA: Washington State Institute for Criminal Justice.