Research Brief
Content Validity of the Static Risk Individual Need Guide for Recidivism (STRONG-R)

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

WASHINGTON STATE
UNIVERSITY

Content Validity[1] of STRONG-R

"*Content validity is the degree to which elements of an assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose* (Haynes, Richard & Kubany, 1995, p. 238). Risk-needs assessment instruments are able to accurately measure the theoretically relevant risk and needs factors when they are supported by validity evidence on the basis of test content (Sireci & Faulkner-Bond, 2014). No assessment instrument is inherently valid unless one gives a purpose for the test (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999). With a large pool of items, Static Risk Individual Need Guide – Revised (STRONG-R) is fully customizable to meet the needs of agencies and jurisdictions in the Washington State to complete a wide range of tasks at each stage of the system, such as failure to appear [FTA], infractions, recidivism, treatment need, and technical violations. Furthermore, the STRONG-R can accurately measure a number of individuals' social and psychological risk and needs factors to allow treatment and intervention to be customized based on the level of individuals' risk and needs. The current study summarized the content validity of the STRONG-R with the purpose of accurately measuring individuals risk and needs[2].

Method

When evaluating content validity of an instrument, four aspects of evidence need to be addressed: domain definition, domain representation, domain relevance, and appropriateness of the test development process (Sireci & Faulkner-Bond, 2014). Both quantitative and qualitative evidence were also used for content validation purpose; in other words, experts' judgment, logical reasoning, and empirical analysis are all qualifying evidence to evaluate content validity of a given test (AERA et. al., 1999).

Content validation has a dynamic nature, meaning it is ongoing assessment process in which psychometric findings and refining conceptual definition are complementary (Riley et. al, 2010). To date, STRONG-R has been assessed on a series of validity tests (Mei & Hamilton, 2016a; Mei & Hamilton, 2016b; Mei & Hamilton, 2016c; Mei &Routh & Hamilton, 2016). Currently, the STRONG-R has passed the first developmental phase, and the content of the test is currently being refined on the basis empirical and psychometric findings. Although further studies are needed for additional improvement, we believe we have amassed a considerable amount of evidence to support the content validity of the STRONG-R.

In the following section, we will address each of the four aspects of the content validity evidence. First, we describe the test construction procedures that were assessed by measurement and subject matter experts (SMEs). Second, we defined the constructs, including scales and sub-scales based on current psychometric evidence. Third, we identified and grouped the measurement items on the STRONG-R to illustrate how items represent and measure their respective scale and sub-scales. Fourth, we examined individual loadings to assess each item's measurement of the constructs and scales.

---

[1] "Validity based on test content" and "content validity" has been used interchangeable.
[2] The risk and needs factors were identified and confirmed via a series psychometric analyses (see Mei & Hamilton, 2016).

Results

1. Appropriateness of the Test Development Process

To support content validity, the test construction and development procedure must be subjected to strong quality control procedures (Sireci & Faulkner-Bond, 2014; Downing & Haladyna, 2006; Holland & Wainer, 1993). A Comprehensive Strategy and Implementation Plan (CSIP) was proposed and implemented in Washington in 2014 (Rist, Hosman & Hamilton, 2014). The CSIP utilized a development team, consisting of a "diagonal slice" of 45 the Department's line staff, using SMEs from several areas of the agency (i.e., prisons, community, and health services). These SMEs provided feedback regarding the technical accuracy of content and the sensitivity of responses. In addition, measurement experts (Barnoski and Lee) reviewed the tool's performance using a standard set of psychometric principles.

Training and pilot testing of the assessment began in 2015 (Hamilton, Kigerl & Routh, 2016). The pilot group of 45 WADOC staff members was trained to administer the STRONG-R and completed assessments using a random sample of 200 individuals. A survey was also distributed and analyzed providing further insight with regard to the STRONG-R's implementation efforts. These findings identified items and responses with differential functioning and difficulty. Further examinations of items and functioning among select sub-groups (i.e., youthful/aging, prison, community, and sex individuals) were outlined for further inspection and modification prior to the full implementation of the tool.

2. Construct Identification and Confirmation – Psychometric Evidence

In previous studies, using psychometric analytic strategies[3] and item selection criterions[4], we identified the STRONG-R's internal latent structure (Mei & Hamilton, 2016), assessed the reliability of its scales (Mei & Hamilton, 2016), examined its concurrent validity (Mei & Hamilton, 2016), and evaluated its scales' convergent and discriminate validity (Mei, Routh & Hamilton, 2016). These tests resulted in the identification of five constructs were identified that draw strong connections with criminological, social and psychological theories and empirical evidence. These five constructs are *(1) Anti-Social History, (2) Education & Employment, (3) Anti-Social Propensity, (4) Substance Abuse,* and *(5) Reintegration Needs.* In total, these five constructs are the measrues of one's Global Risk-Needs, or the propensity of recidivism overall. Furthermore, each of the identified five constructs further includes scales and sub-scales. Based on the content for each construct, scale and sub-scale, each domain was given an operational definition.

---

[3] We have used multiple psychometric analytic techniques, including Exploratory Factor Analysis (EFA), Confirmatory Factor Analysis (CFA), High Order Modeling, Multiple Group Confirmatory Factor Analysis (MGCFA), item invariance tests and scale invariance tests.
[4] "Items are selected/removed include local dependence, differential item functioning, inadequate uni-dimensionality, lack of monotonicity and poor IRT model fit" (Riley et. al, 2010, p. 1312).

Table 1. Operational Definition – Anti-social History

{Construct}/[scale]/(Sub-scale)	Operational Definition
{Anti-social History}	is the extent of individuals' engagement in prior anti-social behavior.
[Criminal History]	is the extent of individuals' involvement with criminal justice system.
(Adult Anti-Social Record)	is the extent of an adult individuals' involvement in criminal justice system.
(Prison Infraction Record)	is the extent of an individual's anti-social behaviors during incarceration.
[Violent History]	is the extent of violence demonstrated in an individual's criminal record.
(Domestic Violence & Violent Misdemeanor History)	is extent of an individual's criminal convictions for initial and more common forms of violence.
(Aggression History)	is the extent to which an individual past behavior indicates he or she is capable of more serious and less common violent convictions.

3. Operational Definition

Domain definition is a critical component of content validity, and it provides information with regard to the intended measures of the construct. A test that has good content validity transforms the abstract notion of the construct to more tangible content domains (Sireci & Faulkner-Bond, 2014). A clear domain definition acknowledges the measureable aspects of the construct. Hence, on the basis of the internal structure identified (Mei & Hamilton, 2016), we refined the operational definitions of the STRONG-R's constructs, scales and sub-scales.

As presented in Table 1, the construct **Anti-social History** is the extent of individuals' engagement in prior anti-social behavior. Under the construct Anti-social History, we have identified two scales, namely, Adult Criminal History and Violent History.

Table 2. Operational Definition – Education/Employment

{Construct}/[Sub-scale]/(Sub-sub-scale)	Operational Definition
{Education & Employment}	is the extent to which an individual's education and employment experience could provide legal income to support pro-social lifestyle.
[Education & Working History]	assesses the individuals' degree of performance in the aspects of work, education, and finance.
(Education Level)	assesses the level of education an individual achieved.
(Working Experience)	assesses an individual's employment history and working skills.
(Juvenile Record)	is an individual's level of involvement in criminal justice system during adolescence diverted him or her from education and employment.
[Income Source]	identifies the legal/illegal sources of an individual's income.
(Illegal Income)	is the extent to which an individual's finical income comes from illegal sources.
(Legal Income)	assesses the level of and individual's legal income and legal employment status.

Criminal History is the extent of individuals' involvement with criminal justice system. It contains two sub-scales, *Adult Criminal Record* and *Prison Infraction Record*. *Adult Criminal History* is the extent of an adult individuals' involvement in criminal justice system. *Prison Infraction Record* is the extent of an individual's anti-social behaviors during incarceration. *Violence History* is the extent of violence demonstrated in an individual's criminal record. It contains two sub-scales, *Domestic Violence and Misdemeanor* and *Aggression History*. *Domestic Violence and Misdemeanor* is extent of an individual's criminal convictions for initial and more common forms of violence. *Aggression History* is the extent to which an individual past behavior indicates he or she is capable of more serious and less common violent convictions.

As showed in Table 2, the construct **Education and Employment** is the extent to which an individual's education and employment experience could provide legal income to support pro-social lifestyle. Education and Employment contains two scales, ***Education & Working History*** and ***Income Source***.

Education and Working Experience assesses the individuals' degree of performance in the aspects of work, education, and finance. It contains two sub-scales, *Education Level* and *Working Experience* as well as *Juvenile Record*. *Education Level* assesses the level of education an individual achieved. *Working Experience* assesses and individual's employment history and working skills. *Juvenile Record* is an individual's level of involvement in criminal justice system during adolescence diverted him or her from education and employment

Income Source identified the legal/illegal sources of an individual's income, and it contains two sub-scales, *Illegal Income* and *Legal Income*. *Illegal Income* is the extent to which an individual's finical income comes from illegal sources. *Legal Income* assesses the level of and individual's legal income and legal employment status.

As displayed in Table 3, the construct **Anti-Social Propensity** assesses the extent to which an individual is likely to engage in anti-social behaviors. It contains four scales, ***Anti-social Influence***, ***Anti-social Personality***, ***Anti-social Cognition*** and ***Violence Propensity***.

Anti-social Influence assesses the extent to which an individual is associated with significant others who demonstrated anti-social behavioral patterns. It contains three sub-scales, *Friend Anti-social, Partner Anti-social* and *Family Anti-social*. *Friend Anti-social* is the degree to which an individual associates with anti-social friends; *Partner Anti-social* is the degree to which an individual associates with an anti-social partner; *Family Anti-social* is the degree to which an individual associates with anti-social family members.

Anti-social Personality is the degree to which an individual demonstrates anti-social behavioral pattern. It contains four sub-scales, *Deception*, *Empathy*, *Retreat* and *Egocentrism*. *Deception* assesses the degree of an individual truthfulness and credibility; *Empathy* is the extent to which a person lacks the capability to care about others and care about morality; *Retreat* assesses to what extent a person has retreated from a pro-social and productive lifestyle; *Egocentrism* is the degree to which a person needs satisfaction, gratification and thrills from different sources, including engaging in diverse types of criminal behaviors.

Anti-social Cognition assesses an individual's attitudes, values, briefs and rationalizations that are favorable to crime. It contains four sub-scales *Respect*, *Change*, *Thinking Errors* and *Cognitive Skills*. *Respect* assesses the amount of respect an individual displays toward authority and property; *Change* assesses an individual's readiness to change from pro-criminal to pro-social lifestyle; *Thinking Errors* assesses the degree to which an individual does not engage in consequential thinking and impulsivity. *Cognitive Skills* assesses the extent to which an individual uses pro-social cognitive skills when dealing with others and internalized problems.

Table 3. Operational Definition – Anti-Social Propensity

{**Construct**}/[*Sub-scale*]/(Sub-sub-scale)	**Operational Definition**
{**Anti-Social Propensity**}	assesses the extent to which an individual is likely to engage in anti-social behaviors.
[**Anti-social Influence**]	assesses the extent to which an individual is associated with significant others who demonstrated anti-social behavioral patterns.
(Friend Anti-social)	is the degree to which an individual associates with anti-social friends.
(Partner Anti-social)	is the degree to which an individual associates with an anti-social partner.
(Family Anti-social)	is the degree to which an individual associates with anti-social family members.
[**Anti-social Personality**]	is the degree to which an individual demonstrates anti-social behavioral pattern.
(Deception)	assesses the degree of a person's truthfulness and credibility.
(Empathy)	is the extent to which a person lacks the capability to care about others and care about morality.
(Retreat)	assesses to what extent a person has retreated from a pro-social and productive lifestyle.
(Egocentrism)	is the degree to which a person needs satisfaction, gratification and thrills from different sources, including engaging in diverse types of criminal behaviors.
[**Anti-social Cognition**]	assesses an individual's attitudes, values, briefs and rationalizations that are favorable to crime.
(Respect)	assesses the amount of respect an individual displays toward authority and property.
(Change)	assesses an individual's readiness to change from pro-criminal to pro-social lifestyle.
(Thinking Errors)	assesses the degree to which an individual does not engage in consequential thinking and impulsivity.
(Cognitive Skills)	assesses the extent to which an individual uses pro-social cognitive skills when dealing with others and internalized problems.
[**Violence Propensity**]	assesses an individual's aggression issues and relative levels of aggressive behaviors/attitudes.
(Physical Abusive)	assesses an individual's level of violence exerted on vulnerable population and animals.
(Hyper-Masculinity)	assesses and individual's violence level toward populations that are capable of effective defense.
(Destructive)	assesses and individual's level of aggression toward property.
(Fixation)	assesses an individual's level of threats and/or obsession with an unhealthy relationship or a threating level of obsession.
(Irritability)	assesses the ease at which an individual can be agitated.

Violence Propensity assesses an individual's aggression issues and relative levels of aggressive behaviors/attitudes. It contains five sub-scales: *Physical Abusive*, *Hyper-Masculinity*, *Destructive*, *Fixation* and *Irritability*. *Physical Abusive* assesses an individual's level of violence exerted on vulnerable population and animals. *Hyper-Masculinity* assesses and individual's violence level toward populations that are capable of effective defense. *Destructive* assesses and individual's level of aggression toward property. *Fixation* assesses an individual's level of threats and/or obsession with an unhealthy relationship or a threating level of obsession. *Irritability* assesses the ease at which an individual can be agitated.

As presented in Table 4, the construct **Substance Abuse** contains three scales, *Hard Drug Use*, *Use-Share-Exchange*, and *Drug Related Crime*. *Hard Drug Use* assesses the extent of which individual's drug addiction problem; *Use-Sharing-Exchange* assesses the extent to which an individual's substance use is collective, or used with associates; *Drug Related Crime* assesses to what extent the individual is involved in drug-related crimes.

Table 4. Operational Definition – Substance Abuse

{**Construct**}/[*Sub-scale*]/(Sub-sub-scale)	**Operational Definition**
[Substance Abuse]	is the degree to which an individual's susceptibility to engage in drug related crime and drug defined crimes.
(Hard Drug Use)	assesses the extent of which individual's drug addiction problem.
(Use-Sharing-Exchange)	assesses the extent to which an individual's substance use is collective, or used with associates.
(Drug Related Crime)	assesses to what extent the individual is involved in drug-related crimes.

As showed in Table 5, the construct **Reintegration Needs** assesses the degree of obstacles an individual has to overcome in order to be successfully reintegrated into society. It contains three scales, *Mental Health-Suicidal*, *Employment Barrier* and *Reentry Needs*.

Table 5. Operational Definition – Reintegration Needs

{**Construct**}/[*Sub-scale*]/(Sub-sub-scale)	**Operational Definition**
{**Reintegration Needs**}	assesses the degree of obstacles an individual has to overcome in order to be successfully reintegrated into society.
[Mental Health –Suicidal]	is the degree to which an individual's mental health condition.
(Suicidal Propensity – History)	assesses one's suicidal tendency in their lifetime.
(Suicidal Propensity – Recent)	assesses one's suicidal tendencies in the most recent 6 months.
[Employment Barrier]	is the extent of obstacles an individual is to overcome in order to have or keep his or her job.
(Physical & Mental Barrier)	assesses the degree to which a person mental and psychical health conditions block his or her employment opportunities.
(Necessity)	assesses to what extent an individual possesses skills and motivation to work.
(Work Ethic)	assesses positive individual employment experiences.
(Systematic Barrier)	assesses obstacles that cannot be easily overcome by oneself.
[Reentry Needs]	assesses the degree to which an individual needs additional help from social and justice system to meet daily, basic needs.

Reentry Needs assesses the degree to which an individual needs additional help from social and justice system to meet daily and basic needs, such as the need of housing, food and cloth.

Mental Health-Suicidal is the degree to which an individual's mental health condition. It contains two sub-scales, *Suicidal Propensity – History* and *Suicidal Propensity – Recent*. *Suicidal Propensity – History* assesses one's suicidal tendency in their lifetime; *Suicidal Propensity – Recent* assesses one's suicidal tendencies in the most recent 6 months.

Employment Barrier is the extent of obstacles an individual is to overcome in order to have or keep his or her job. It contains four sub-scales, *Physical & Mental Barrier*, *Necessity*, *Work Ethic* and *Systematic Barrier*. *Physical & Mental Barrier* assesses the degree to which a

person mental and psychical health conditions block his or her employment opportunities; *Necessity* assesses to what extent an individual possesses skills and motivation to work; *Work Ethic* assesses positive individual employment experiences; *Systematic Barrier* assesses obstacles that cannot be easily overcome by oneself.

3. Measures of Scales - Domain Relevance and Domain Representation

Along with domain definition, domain representation and domain representation are major criterions for content validity. Domain representation concerns whether a construct's domains are adequately represented in the assessment instrument, while domain relevance assesses to what extent the items on the assessment instrument is relevant to the corresponding domains. Traditionally, content validation involves subject-matter-experts (SMEs)' subjective judgment on the extent to which the items are relevant to and can represent the targeted constructs (Messick, 1993; Hayens, Richard & Kubany, 1995; Sireci & Faulkner-Bond, 2014).

Although, currently, a content validation study has been scheduled in which SEMs' input will be used to improve the content validity of the identified constructs, to date, we have found substantial objective evidence to support the content validity of the STRONG-R. Because the traditional method of soliciting experts' subjective judgment such as rating the relevance of the items to the targeted constructs is equivalent to expert's perception regarding correlation al relationship between the items and the targeted construct (Gajewski et. al., 2011). We believe that the objective correlational relationships between of item and the targeted construct are empirical evidence to support domain relevance and domain representativeness. Therefore, on the basis of the psychometric analytic evidence (Mei & Hamilton, 2016a; Mei & Hamilton, 2016b; Mei & Hamilton, 2016c; Mei, Routh & Hamilton, 2016), we objectively quantified the relationship between the items and the constructs, as indicated by an estimated model parameter, namely, the standardized loading. We argue that if an item is relevant to and representative of the targeted construct, it should demonstrate a substantial standardized loading value. Likewise, if a lower order factor is relevant to and representative of the higher order factor, then it should demonstrate a substantial standardized loading value for the representativeness and relevance of the targeted construct.

First, as presented in Table 6, the item loading values for the first order/subscales ranged from .671 to .868. Second, the first order factor loadings ranged from .726 to .894. Third, the second order factor loading ranged from .720 to .848. Last but not least, the construct loading ranged from .899 to .984. With these statistically significant and substantial loading at all levels within their internal structure, we contend that the items and the identified latent factors at higher level in the STRONG-R demonstrated their domain relevance and domain representativeness.

Table 6. Domain Relevance and Domain Representation

3rd Order Factor	3rd Order Factor Loading	2nd Order Factor	2nd Order Factor Loading	1st Order Factor	1st Order Factor Loading	1st Order Item Loading	No. Of Items
Anti-Social History	.950	Criminal History	.721	Adult Criminal Record	.894	.868	3
				Prison Infraction Record	.894	.868	4
		Violence History	.721	DV & Violent Misd. Hx.	.894	.868	4
				Aggression Hx.	.894	.868	3
Education & Employment	.967	Education & Work Experience	.848	Juvenile Record	.848	.714	4
				Education Level	.848	.714	2
				Work Experience	.848	.714	2
		Income	.848	Illegal Income	.848	.714	2
				Legal Income	.848	.714	2
Anti-Social Propensity	.984	Anti-social Influence	.720	Friend Anti-social	.877	.789	3
				Partner Anti-social	.877	.789	3
				Family Anti-social	.877	.789	3
		Anti-social Personality	.720	Deceit	.877	.789	2
				Empathy	.877	.789	2
				Retreat	.877	.789	3
				Egocentrism	.877	.789	7
		Anti-social Cognition	.720	Respect	.877	.789	2
				Change	.877	.789	2
				Thinking Error	.877	.789	2
				Cognitive Skills	.877	.789	2
		Violence Propensity	.720	Physical Abusive	.877	.789	3
				Hyper-Masculinity	.877	.789	3
				Destructive	.877	.789	2
				Fixiation	.877	.789	3
				Violence Out of Control	.877	.789	4
Substance Abuse	.899	--	--	Hard Drug Use	.927	.671	3
				Drug Use-Share-Barter	.927	.671	3
				Drug Related Crime	.927	.671	3
Reintegration Needs	.924	Employment Barrier	.727	Physical & Mental Barrier	.726	.845	2
				Social & Working Skill	.726	.845	3
				Work Habit	.726	.845	3
				Systematic Barrier	.726	.845	3
		Mental Health	.727	Mental Health Issue - Past	.726	.845	3
				Mental Health Issue - Recent	.726	.845	3
		Reentry needs	.727	--	--	.845	9

Conclusion

Given the dynamic nature of content validity, both objective quantitative psychometric evidence and subjective SMEs' judgments are essential component of the iterative content validation process. Based on the available psychometric evidence and SMEs' involvement in the development process, we believe that a considerable amount of quantitative and qualitative evidence has been amassed to support the content validity of the STRONG-R for the purpose of assessing a wide range of individual's risk and needs factors that are associated with recidivism. Yet, currently, the aspects of the content validity of the STRONG-R will be further examined by using SMEs' subjective judgment and future empirical evidence.

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.

Downing, S., & Haladyna, Thomas M. (2006). *Handbook of test development*. Mahwah, N.J.: L. Erlbaum.

Gajewski, B., Coffland, V., Boyle, D., Bott, M., Price, L., Leopold, J., & Dunton, N. (2012). Assessing Content Validity Through Correlation and Relevance Tools: A Bayesian Randomized Equivalence Experiment. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences, 8*(3), 81-96.

Hamilton, Z., Kigerl, A., & Routh, D. (2016). *The STRONG-R pilot assessment study*. Spokane, WA: Washington State Institute for Criminal Justice.

Haynes, S. N., Richard, D. & Kubany, E. (1995). Content Validity in Psychological Assessment: A Functional Approach to Concepts and Methods. *Psychological Assessment, 7*(3), 238-47.

Holland, P., Wainer, H, & Educational Testing Service. (1993). *Differential item functioning*. Hillsdale: Lawrence Erlbaum Associates.

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

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

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

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

Riley, W., Rothrock, T., Bruce, N., Christodolou, B., Cook, C., Hahn, K., & Cella, E. (2010). Patient-reported outcomes measurement information system (PROMIS) domain names and definitions revisions: Further evaluation of content validity in IRT-derived item banks. *Quality of Life Research, 19*(9), 1311-1321.

Rist, M., Hosman, S., & Hamilton, Z. (2014). *STRONG-R Individual Change: Comprehensive Strategy and Implementation Plan*. Olympia, WA: Washington State Department of Corrections.

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

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

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