

---

# COMPAS Scales and Risk Models Validity and Reliability

---

A SUMMARY OF RESULTS  
FROM  
INTERNAL AND INDEPENDENT STUDIES

RESEARCH AND DEVELOPMENT DEPARTMENT

JULY 20, 2010



# Contents

<b>Table of Contents</b>	<b>i</b>
<b>1 COMPAS Validity and Reliability</b>	<b>2</b>
1.1 Test-Retest Reliability . . . . .	3
1.2 Predictive Validity . . . . .	3
1.3 Construct Validity . . . . .	7
1.4 Internal Consistency Reliability . . . . .	10
1.5 Validity of our specific Needs Scales . . . . .	10
1.5.1 The issue of Content Validity . . . . .	12
<b>Bibliography</b>	<b>14</b>

# Chapter 1

## COMPAS Validity and Reliability

The COMPAS is a fourth generation risk and needs assessment instrument. It is a computerized tool designed to assess offenders' needs and risk of recidivism and is used by criminal justice agencies across the nation to inform decisions regarding the placement, supervision, and case management of offenders. The instrument was empirically developed and focuses on predictors known to affect recidivism. It includes dynamic risk factors in its prediction of recidivism, and provides information on a variety of well validated risks and need factors designed to aid in correctional treatment to decrease the likelihood that offenders will re-offend.

In this section we summarize research findings from multiple studies that demonstrate COMPAS is reliable (Test-Retest and Internal Consistency), that its scales measuring needs have construct validity and behave consistently and finally we summarize studies of the predictive validity of the COMPAS risk-scales. An overall conclusion is that COMPAS was found to be reliable and has good predictive and construct validity.

We acknowledge at the outset that much of the evidence for the reliability and validity of the COMPAS is found in the results of research studies conducted by Northpointe. We know that critics may discount this research. However, most of our in-house research is conducted for state agencies, and that competent research divisions within those agencies closely scrutinize our methods and results. Such state-sponsored studies are, thus, often subjected to a far more thorough vetting than that provided by the editors of peer-reviewed journals often resulting from the fact that such agencies have direct access to the same data, can scrutinize such data and often can replicate and test our findings.

We recognize that independent research evidence for the reliability and validity of the COMPAS will bolster its standing in the marketplace. However, the

lack of independent research is partly a function of data access. We have consistently encouraged our clients to form collaborative relationships with independent researchers to pursue independent research opportunities and conduct well-designed validation studies.

Furthermore, Northpointe has an established history of working in partnership with our clients to advance knowledge and practice. From our early work in jail classification to our recent partnership with CDCR and the University of Cincinnati, Northpointe leverages the opportunity of public and private partnership to test and advance knowledge. The results are shared both in peer-reviewed publications, multiple presentations at professional conferences, and with detailed technical reports to clients. Thus, our research and evaluation findings are publicly shared in technical reports, peer-reviewed articles and book chapters to advance the availability of current information for use in practice.

## 1.1 Test-Retest Reliability

This is one of the key aspects of reliability. In a recent independent study by Farabee, Zhang, Roberts, and Yang (2010) the COMPAS scales showed very high test-retest reliability, with correlations ranging from .70 to 1.00, and with an average across all its scales over 0.80. Thus, the various COMPAS sub-scales demonstrated good to excellent reliability over time. An important aspect of the Farabee study was a comparison against the well known LSI (the industry leader). Overall, the average test-retest correlation coefficient for the COMPAS scales was .88; for LSI-R, the mean as measured in the same study was .64. These results indicated a distinct superiority of the COMPAS compared to the LSI scales.

## 1.2 Predictive Validity

The COMPAS distinguishes between risk scales (designed to predict recidivism) and needs scales (designed to measure needs, inform case plans and identify intervention targets). This approach of separating risk and needs aligns with current best practices in risk assessment (Baird, 2009; Gottfredson & Moriarty, 2006)

The COMPAS has two main risk models: General Recidivism Risk and Violent Recidivism Risk. Again, following best practices, we believe risk scales designed to predict risk should be dynamic (composed of dynamic, criminogenic needs) so that one can measure changes in risk of recidivism over time.

Others have argued that risk models should be composed of static criminal history factors available in criminal justice information management systems, arguing that these models are more objective, reliable, and efficient (Barnoski & Drake, 2007).

The current COMPAS Recidivism Risk scale is an equation derived from a regression model developed in a sample of presentence investigation and probation intake cases in 2002. It was trained to predict any offense arrest within two years of intake assessment. This risk scale is composed of COMPAS items selected through diagnostic modeling strategies. Whereas scale reliability and coherence were emphasized for developing the other COMPAS scales, predictive validity was prioritized in developing this new recidivism scale. The scale is described below and thus aims to predict which offenders will commit crimes subsequent to their initial COMPAS screening date. Our methods for developing and validating these scales were strongly influenced by the research of John Copas and colleagues who have developed an outcomes-based recidivism scale for England and Wales (Copas & Marshall, 1998). Some of these methods are described in recent books by Hosmer and Lemeshow (2000) and Harrell (2001). Standard logistic regression was used to predict recidivism with the full set of variables in each candidate set. We also used bootstrap validation to generate highly realistic estimates of predictive validity (Harrell, 2001; le Cessie & van Houwelingen, 1992). The model on which the Recidivism Risk scale is based was selected from several competing models on the basis of parsimony and predictive power.

The second major COMPAS scale is the Violent Recidivism Risk scale. It is also an equation derived from a regression model and was developed in 2006 in a sample of presentence investigation and probation intake cases. It was trained to predict violent offenses (misdemeanor and felony) within two years of intake.

Northpointe is committed to testing, evaluating, and improving our risk models. During initial phases of scale development we rely on alternative methods to test the early stage criterion validity of the risk models, and we also include well known criterion factors such as age-at-first arrest to ensure that fundamental associations are present. The above General Recidivism Risk and Violent Recidivism Risk scales have also subsequently been validated with prospective outcomes in new samples in several different studies since they were first developed.

Repeated validation studies across jurisdictions and agencies: We typically include an outcomes study in most pilot tests of the COMPAS to evaluate the predictive validity of the risk scales in each new jurisdiction. In 2006 we conducted pilot tests in the New York Division of Probation and Correctional Alternatives (DPCA), the New York State Division of Parole (NYSDP), and

the Michigan Department of Corrections (MDOC). These three pilot studies all had outcomes studies built into them. In 2008 we conducted outcomes studies at all three sites using their pilot data. We also conducted separate studies in the California Department of Corrections and Rehabilitation (CDCR) and DPCA. This latter study was recently published in the *Journal of Criminal Justice and Behavior* (Brennan, Dieterich, & Ehret, 2009).

Table 1.1 shows the results of tests of the predictive validity of the COMPAS risk scales from the outcomes studies conducted in the last three years (Brennan & Dieterich, 2007, 2008b, 2008a; Brennan et al., 2009; Brennan & Dieterich, 2009; Dieterich, Brennan, & Breitenbach, 2010; Farabee et al., 2010). The table shows the Area Under the Curve (AUC) for the General Recidivism Risk and Violent Recidivism Risk scales. The AUC is the most widely used measure of predictive accuracy in criminal justice, psychology, medicine, and related fields. AUCs of .65 to .69 indicate modest to moderate predictive accuracy. AUC's of .70 to .75 indicate moderate to strong predictive accuracy. We also evaluate our risk scales using other criteria, including failure probabilities, odds ratios, and hazard ratios. Note that for felony arrest, abscond, and return outcomes, the Recidivism Risk Scale is tested. For person offense arrests the Violent Recidivism Risk Scale is tested. The results of these studies indicate that the COMPAS risk scales generally fall into the modest to strong range of predictive accuracy. They also indicate that COMPAS generally meets or exceeds the AUC values produced by competitive instruments such as the LSI. In some of the studies conviction data was used, because arrest data was unavailable (Dieterich et al., 2010).

In the UCLA study the outcomes were based on the DOJ arrest data obtained by the authors (Farabee et al., 2010). Note that the criterion used to test the Violent Recidivism Risk Scale in this study was a binary measure indicating if the first arrest within two years of release was for a person offense. If the first offense was for a nonviolent offense, the study treated the case as a non-recidivist, even if that case subsequently committed a violent offense during the follow-up.

*An independent but flawed study of Predictive Accuracy:* A recent independent study by Fass, Heilbrun, DeMatteo, and Fretz (2008) purporting to examine the predictive validity of the COMPAS risk scales concluded that its effectiveness varies widely by ethnicity. This conclusion was subsequently disputed in a peer-reviewed paper (Brennan et al., 2009). The latter noted that the Fass study only had outcome data available on a relatively small sample of 276 subjects (not their full sample) and that this sampling issue was a serious flaw that undermines their conclusions. This sampling issue appears to have been missed by the journal reviewers. The glaring problem is that when Fass broke down their small sample by ethnicity the resulting ethnic specific sub-samples

Table 1.1: Summary of AUC results for Recidivism Risk and Violent Recidivism Risk Scales in several outcomes studies.

Study	N	Year	Any				Return
			Arrest	Felony	Person	Abscond	
CDCR Study	(n=20,890)	2007					.672
NYSDP Pilot	(n=553)	2008	.679		.728	.652	
DPCA Pilot	(n=987)	2009	.730				
DPCA Study	(n=2,328)	2009	.707	.717	.742		
MDOC Study <sup>a</sup>	(n=9,685)	2010	.720	.710	.720	.690	
UCLA Study	(n=25,009)	2010	.700		.610 <sup>b</sup>		

<sup>a</sup> Study used conviction data.

<sup>b</sup> Flawed outcome criterion, see text.

essentially became too small to properly or reliably measure a predictive effect. For example, there was a grand total of only 1 failure (recidivism) in the subsample of Caucasian subjects – thus the effective sample size for the analysis of Caucasians was 1. To base any serious conclusion on such a sample can hardly be seen as conclusive or reliable evidence. In our own studies for MDOC we found no significant differences in the predictive validity for different ethnic groups (Brennan et al., 2009).

*Comparing COMPAS predictive accuracies against other established instruments:* The AUC's of the other main instruments often used for offender risk prediction may help to contextualize the findings of our studies. Perhaps the best known instruments are the Violence Risk Appraisal Guide [VRAG] (Quinsey, Harris, Rice, & Cormier, 1998); the Level of Services Inventory-Revised [LSI-R] (Andrews & Bonta, 1995); and the Psychopathy Checklist-Revised [PCL-R] (Hare, 1991). The AUC values for these instruments in recent studies are quite varied according to the specific populations, outcome periods, and dependent variables used in specific studies.

Regarding the VRAG, Quinsey et al. (1998) found an AUC of 0.76 in a large scale, multiyear recidivism study. In Barbaree, Seto, Langton, and Peacock (2001) reported AUCs of 0.69 in predicting serious re-offending and 0.77 when predicting any re-offense for sex offenders. The study of Kroner, Stadtland, Eidt, and Nedopil (2007) obtained an AUC of 0.703 in a study of re-offending among mentally ill offenders.

Regarding the LSI-R, the recent review by Andrews, Bonta, and Wormith (2006) did not provide AUCs. However, Barnoski and Aos (2003) found AUCs of 0.64-0.66 for the LSI-R in predicting felony and violent recidivism among Washington State prisoners. The study of Flores, Lowenkamp, Smith, and Latessa (2006) report an AUC of 0.689 using the LSI-R to predict re-

incarceration among federal probationers. Dahle (2006) reported an AUC of 0.65 using the LSI-R to predict violent recidivism. Barnoski (2006) reported an AUC of 0.65 using the LSI-R to predict felony sex recidivism.

PCL-R: AUC levels again varied across studies. For example, in a Swedish study of mentally ill violent offenders, Grann, Belfrage, and Tengstrom (2000) found AUC levels of 0.64-0.75 based on various follow-up timeframes. Barbaree et al. (2001) reported AUCs of 0.61, 0.65, and 0.71 for the PCL-R in predicting various recidivism outcomes among sex offenders.

### 1.3 Construct Validity

Construct validity refers to observed correlations between measures of the same or divergent constructs. Construct validity is relevant only for the COMPAS needs scales and refers in part to unidimensionality of the scale and to its factor structure. It additionally is based on establishing evidence that it correlates in an expected manner with similar scales, and to other relevant scales in theoretically expected ways.

Viewing the LSI-R as a gold-standard since it is the current industry leader, a direct approach to construct validity is to measure the correlation between matched scales of the LSI and COMPAS. This would be a good indication for how well the COMPAS scales are measuring the same related concept. Thus, in the following we tabulate the correlations between the LSI-R and COMPAS scales as found in the report by Farabee et al. (2010). The results of this analysis show a direct and high level of correlation between the matching scales. This offers very strong evidence of the construct validity of the COMPAS scales, i.e., that they are measuring largely similar or overlapping constructs to those criminogenic constructs as assessed by the LSI. However, it should be noted that COMPAS contains a number of additional scales that are omitted from the LSI-R, and also that in general, the reliabilities of the COMPAS scales are, in almost all cases, superior to those of the LSI.

Shifting to more general issues of construct validity we continue to accumulate a variety of additional evidence to support the construct validity of the COMPAS need scales. For example the COMPAS substance abuse measure correlates positively (.44) with the SASSI in our MDOC pilot sample, and as seen from the above table .53 with the LSI drug scale.

Construct validity is also demonstrated if a measure correlates in the predicted manner with other variables with which it theoretically should correlate. For example, research in developmental delinquency (longitudinal research in which anti-social behaviors and attitudes are studied over the life course) consistently finds that youth with early onset of delinquent behavior

Table 1.2: Correlations between COMPAS and LSI-R scales

COMPAS	LSI-R	Correlation
Criminal Involvement	Criminal History	.64 ( $p < .0001$ )
Criminal Associates/Peers	Companions	.48 ( $p < .0001$ )
Substance Abuse	Alcohol/Drug Problem	.53 ( $p < .0001$ )
Financial	Financial	.49 ( $p < .0001$ )
Vocation/Education	Education/Employment	.51 ( $p < .0001$ )
Family Criminality	Family/Marital	.16 ( $p > .10$ )
Leisure	Leisure/Recreation	.05 ( $p > .10$ )
Residential Instability	Accommodation	.57 ( $p < .0001$ )
Criminal Attitudes	Attitudes/Orientation	.20 ( $p = .08$ )

tend to have more serious delinquency trajectories and more negative emotionality, lower achievement, and problems in social adjustment (Moffitt, 1993). Thus, when we consistently find, over multiple studies, that our Criminal Personality, Criminal Attitudes, Social Adjustment and Vocational Educational scales correlate with age-at-first-arrest, just as developmental delinquency research predicts, we take this as evidence of construct validity. Furthermore, age-at-first-arrest is a good external variable to demonstrate construct validity of the COMPAS needs scales. Although age-at-first is collected inside the COMPAS, it comes from official records, while the needs scales are scored using a different method (interview and self-report).

We have evidence of construct validity of this type from psychometric studies in the Michigan Department of Corrections, New York Probation, New York Parole, Georgia Department of Corrections and many other sites. To illustrate our approach to demonstrating construct validity, we present results in 1.3 from a current sample in CDCR. The CDCR sample consists of 6,485 Core COMPAS assessments conducted between September 26, 2008 and January 27, 2009. Men comprise 91% of the sample.

There are many other notable correlation patterns in Table 1.3 that provide evidence of construct validity for the COMPAS scales. For example, we see that age-at-first arrest correlates negatively with the higher-order personality scales Criminal Attitudes ( $p = -.12$ ) and Criminal Personality ( $p = -.15$ ). This comports with findings in developmental research that indicate offenders with early onset are more likely to have high scores on similar types of personality measures and more serious and persistent criminal involvement (Moffitt, 1993). Similarly, we see that offenders with earlier age-at-first arrest are more likely to have higher scores on scales measuring factors that have been identified as criminogenic in longitudinal developmental studies. These scales include Criminal Associates and Peers ( $p = -.28$ ), Family Crime ( $p = -.19$ ), Vocational/Educational Problems ( $p = -.22$ ), and Social Environment ( $p$

	Age-at-First	Prior Arrests	Returns to Custody	Commitments	Assaultive Misconduct
CassPeer	-0.28	0.13	0.17	0.09	0.18
SubAbuse	-0.05	0.23	0.19	0.16	-0.07
Financ	-0.07	0.10	0.11	0.08	0.02
VocEd	-0.22	0.11	0.14	0.06	0.17
FamCrim	-0.19	0.09	0.10	0.05	0.11
SocEnv	-0.18	0.11	0.11	0.10	0.14
Leisure	-0.09	0.10	0.11	0.08	0.08
ResInst	-0.03	0.12	0.15	0.10	0.10
SocAdj	-0.20	0.18	0.19	0.12	0.14
Soc.Isolation	0.04	0.11	0.13	0.10	0.06
CrimAttC	-0.12	0.03	0.05	0.00	0.13
CrimPers	-0.15	0.09	0.13	0.06	0.17

Table 1.3: Correlations of COMPAS Scales with Criminal History Indicators in CDCR Sample.

= - .18) (Farrington, Jolliffe, Loeber, Stouthamer-Loeber, & Kalb, 2001).

Another pattern in Table 1.3 is defined by the correlations between previous arrests and the scales Substance Use ( $p = .23$ ), Financial Problems ( $p = 0.1$ ), Residential Instability ( $p = .12$ ) and Social Isolation ( $p = .11$ ) (Stouthamer-Loeber, Loeber, Wei, Farrington, & Wikstrom, 2002).

There are moderate, significant correlations between the assaultive misconduct item from the COMPAS and the scales Criminal Associates and Peers ( $p = .18$ ), Vocational Educational Problems ( $p = .17$ ), Social Environment ( $p = .14$ ), Social Adjustment ( $p = .14$ ), and Criminal Personality ( $p = .17$ ). In their meta-analysis, Gendreau, Goggin, and Law (1997) found that antisocial attitudes and criminal peers were important individual level predictors of prison misconduct. There are notable, significant correlations between the number of returns to custody for a parole violation and the scales Criminal Associates and Peers ( $p = .17$ ), Substance Abuse ( $p = .19$ ), Vocational Educational Problems ( $p = .14$ ), Residential Instability ( $p = .15$ ), and Social Adjustment ( $p = .19$ ). Substance abuse, residential stability, and employment and education have been identified in past research as some the most important risk and needs factors associated with reentry success (Nelson, Deess, & Allen, 1999).

Overall, the observed relationships between the COMPAS scales and criminal history indicators in the CDCR sample provide evidence of the construct validity of the scales. These correlations comport with relationships between risk factors and serious and violent trajectories observed in developmental crimino-

logical research (Herrenkohl et al., 2000; Tolan & Gorman-Smith, 1998). The significant correlations we have pointed out are somewhat attenuated by variability in the base rates of the paired variables. These modest associations are typical of correlations between risk factors and criminal involvement variables observed in many criminal justice research contexts.

## 1.4 Internal Consistency Reliability

This form of reliability - typically assessed by Cronbach's Alpha Coefficient - is a widely used and popular reliability approach. It is often used as a counterpart to test-retest reliability. Several prior published studies provide data for both the LSI and COMPAS thus allowing direct comparisons of the two instruments. In our own prior reports to CDCR we have provided internal-reliabilities of all the relevant COMPAS scales as assessed on CDCR prisoners. The LSI results for internal reliability, for a variety of correctional samples are published in various journal articles. A comparison of these published findings - mostly using Cronbach's alpha coefficient - also indicates that the COMPAS scales outperformed the LSI scales. Specifically, in a study of  $N = 1,077$  CDCR prisoners the average Cronbach Alpha across the COMPAS scales was 0.7 (Brennan, Dieterich, & Oliver, 2005). An additional COMPAS study of 1,534 San Bernardino probationers had an average Alpha of 0.73 (Brennan, Dieterich, & Oliver, 2006). In contrast an LSI reliability study in a Canadian prison sample had an average Alpha of 0.39 (Simourd, 2004).

We show in Table 1.4 a listing of the Cronbach's Alpha Coefficient for each scale on a large sample consisting of males from CDCR and MDOC.

## 1.5 Validity of our specific Needs Scales

How relevant are our specific needs scales for understanding and predicting criminality? Our specific need scales were selected to measure factors determined through prior research to have an impact on outcomes in criminal justice. The needs scales also are used to guide individualized decisions for case planning, including identifying targets and choosing interventions. Although we view risk scales separately from need scales in terms of function and purpose, both the need scales and the risks scales should be relevant for probation, prison, reentry, and parole work. In other words, while we do not use the need scales to directly predict recidivism, we do want the need scales to measure individual factors such as criminal thinking, education, employment, substance abuse, residential stability and other aspects of the person-in-environment that represent potential relevant targets for interventions.

Table 1.4: Cronbach's Alpha Coefficient for each COMPAS scale

	Items	N	Min	Max	Mean	SD	Alpha
Criminal Peers	7	47679	7.00	22.00	10.81	3.55	0.81
Crim. Attitudes	10	47679	10.00	50.00	22.12	5.55	0.80
Crim. Involvement	4	47679	0.00	19.00	8.25	4.61	0.75
Crim. Opportunity	14	47679	13.00	40.00	21.37	4.71	0.67
Crim. Personality	13	47679	13.00	61.00	30.80	6.45	0.73
Current Violence	7	47679	7.00	13.00	8.18	1.27	0.53
Early Juv. Soc. Fail.	13	47679	7.00	32.00	12.53	4.03	0.69
Family Criminality	6	47679	6.00	12.00	7.39	1.49	0.65
Finance	5	47679	5.00	15.00	8.48	2.51	0.75
History Non-Compl.	5	47679	0.00	21.00	5.98	5.37	0.75
History Violence	9	47679	0.00	22.00	2.68	2.86	0.59
Leisure	5	47679	5.00	17.00	8.22	3.66	0.86
Residential Inst.	10	47679	9.00	31.00	13.72	4.07	0.72
Soc. Isolation	8	47679	8.00	40.00	17.72	5.24	0.81
Soc. Adjustment	15	47679	12.00	38.00	20.48	3.66	0.56
Soc. Environment	6	47679	6.00	12.00	7.86	2.00	0.83
Substance Abuse	10	47679	10.00	20.00	13.10	2.47	0.75
Vocation/Education	11	47679	11.00	30.00	18.85	3.82	0.69

In our psychometric studies we have demonstrated that the need scales measure factors that are relevant for case planning by fitting univariable regression models in which each need scale predicts future recidivism. Table 1.5 shows the results of fitting a survival model to each need scale for predicting a return to prison for a technical violation in the CDCR sample – all are statistically significant. Note that we have similar results/evidence from the other outcomes studies that we have conducted.

The row for Vocational/Education shows the coefficient, hazard ratio, standard error, and t-value from a survival model in which Vocational/Education predicts return to prison for a technical violation. The hazard ratio indicates that for every one-unit increase in the Vocational/Educational decile score, the hazard for return to prison for a technical violation increases by 11%. The contents of the table are sorted by the hazard ratio. Scales with the largest hazard ratio are ranked higher. The top five scales on the basis of hazard ratio are Vocational Educational Problems, Criminal Personality, Social Adjustment, Residential Instability, and Criminal Thinking. If the t-value is greater than 1.96, the effect is significantly different than zero. All these estimates are significant, but also fairly modest in size. The significance level is a function of sample size; the sample consists of 6,485 soon-to-be-released inmates (first release to parole). However, these results demonstrate that the COMPAS need

Table 1.5: Uni-variable Survival Model: Hazard for Return to Prison (Technical Violation)

Scale Decile Score	Coeff	Hazard Ratio	Se (Coeff)	t-value
Vocational/Education	0.101	1.11	0.007	15.47
Criminal Personality	0.081	1.08	0.006	12.66
Social Adjustment	0.076	1.08	0.006	12.45
Residential Instability	0.074	1.08	0.006	12.58
Criminal Thinking	0.057	1.06	0.007	8.59
Leisure and Recreation	0.057	1.06	0.006	9.46
Substance Abuse	0.051	1.05	0.006	7.87
Financial Problems/Poverty	0.048	1.05	0.006	7.87
Social Environment	0.044	1.05	0.006	7.92
Family Criminality	0.043	1.04	0.006	7.12
Social Isolation	0.036	1.04	0.006	5.57

scales measure factors that have clear and predictive relationships to recidivism, and hence, they are useful measures of potential intervention targets.

A natural progression is to move from measuring a need at a single time point (say at intake), to measuring a need at different occasions (completion of treatment). In some cases multiple measurement occasions imply a test of an intervention. This in turn can support experimental or quasi-experimental designs, with a control or comparison group. Testing for intervention effects requires measures that have high validity, reliability, and sensitivity. The measure itself may not always have an effect on the target. The measure is used only to determine how much of the factor exists and if this quantity changes over time as a result of the intervention. Several of the COMPAS scales consist of items that can change over time and thus are dynamic.

### 1.5.1 The issue of Content Validity

Content validity refers to the coverage of key factors that are relevant in the criminogenic domain. It is clear that COMPAS has a more comprehensive coverage of relevant scales than the LSI-R. Content validity has a major role in any assessment field. It refers to the extent to which an assessment comprehensively includes and assesses the key factors in a domain of interest. The LSI includes 10 important criminogenic factors that assess constructs well supported in the literature and that have emerged from over 15 years of meta-analytic research. A study conducted by Farabee et al. (2010) found that 9 out of these 10 LSI scales are clearly matched to a similar scale in COMPAS. Thus, in terms criminogenic scale coverage (content validity), COMPAS

matches virtually all scales contained in the LSI. However, the COMPAS system additionally includes another 14 scales that can be utilized or turned on/off by an agency depending on its information needs.

These additional scales are themselves well supported empirically and include such factors as: anger/hostility, history of non-compliance, low social supports, socialization failure, and so on. In those paired scales between COMPAS and the LSI showing weaker correlations it is also clear that the two instruments are assessing different facets of these constructs. For example, Family Criminality in COMPAS specifically assesses the actual criminality of the family; while Family/Marital in LSI-R focuses on family relationships. We also note that in a study conducted by Farabee et al. (2010) the LSI-R scale for family issues has far lower test-retest reliability than the COMPAS scale (.55 versus .94).

# Bibliography

- Andrews, D. A., & Bonta, J. (1995). *LSI-R: The Level of Service Inventory-Revised*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or needs assessment. *Crime and Delinquency*, 52(1), 7-27.
- Baird, C. (2009). *A question of evidence: A critique of risk assessment models used in the justice system*. (Tech. Rep.). Madison, WI: National Council on Crime and Delinquency.
- Barbaree, H. E., Seto, M., Langton, C. M., & Peacock, E. J. (2001). Evaluating the predictive accuracy of six risk assessment instruments for adult sex offenders. *Criminal Justice and Behavior*, 28(4), 490-521.
- Barnoski, R. (2006). *Sex offender sentencing in Washington State: Predicting recidivism based on the LSI-R*. (Tech. Rep.). Olympia, WA: Washington State Institute for Public Policy. Available from <http://www.wsipp.wa.gov/rtpfiles/06-02-1201.pdf>
- Barnoski, R., & Aos, S. (2003). *Washington's offender accountability act: An analysis of the Department of Corrections' risk assessment (document no. 03-12-1202)* (Tech. Rep.). Olympia, WA: Washington State Institute for Public Policy.
- Barnoski, R., & Drake, E. K. (2007). *Washington's offender accountability act: Department of corrections' static risk assessment* (Tech. Rep.). Olympia, WA: Washington State Institute for Public Policy.
- Brennan, T., & Dieterich, W. (2007). *A test of the predictive validity of the COMPAS risk scales: Predicting return to prison in a large parole sample*. (Tech. Rep.). Northpointe Institute for Public Management.
- Brennan, T., & Dieterich, W. (2008a). *Michigan Department of Corrections Core COMPAS pilot study: One-year follow-up* (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Brennan, T., & Dieterich, W. (2008b). *New York State Division of Parole COMPAS Reentry pilot study: Two-year follow-up*. Traverse City, MI: Northpointe Institute Inc.
- Brennan, T., & Dieterich, W. (2009). *Testing the predictive validity of the*

- DPCA COMPAS risk scales: Phase I* (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Brennan, T., Dieterich, W., & Ehret, B. (2009). Evaluating the predictive validity of the COMPAS risk and needs assessment system. *Criminal Justice and Behavior*, *36*, 21-40.
- Brennan, T., Dieterich, W., & Oliver, B. (2005). The COMPAS scales: Normative data for males and females in community and incarcerated samples [Computer software manual]. Northpointe Institute for Public Management. (Version 4.5)
- Brennan, T., Dieterich, W., & Oliver, W. (2006). *San Bernardino County Probation Department COMPAS psychometric report* (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Copas, J., & Marshall, P. (1998). The offender group reconviction scale: A statistical reconviction score for use by probation officers. *Applied Statistics*, *47*, 159–171.
- Dahle, K. P. (2006). Strengths and limitations of actuarial prediction of criminal re-offence in a German prison sample: A comparative study of LSI-R, HCR-20 and PCL-R. *International Journal of Law and Psychiatry*, *29*(5), 431-442.
- Dieterich, W., Brennan, T., & Breitenbach, M. (2010). *Testing the predictive validity of the Reentry COMPAS risk scales* (Tech. Rep.). Traverse City, MI: Northpointe Institute for Public Management.
- Farabee, D., Zhang, S., Roberts, R. E. L., & Yang, J. (2010). *COMPAS validation study: Final report* (Tech. Rep.). Los Angeles: Semel Institute for Neuroscience and Human Behavior.
- Farrington, D. P., Jolliffe, D., Loeber, R., Stouthamer-Loeber, M., & Kalb, L. M. (2001). The concentration of offenders in families, and family criminality in the prediction of boys' delinquency. *Journal of Adolescence*, *24*, 579–596.
- Fass, T., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS. *Criminal Justice and Behavior*, *35*, 1095–1108.
- Flores, A. W., Lowenkamp, C. T., Smith, P., & Latessa, E. J. (2006). Validating the Level of Service Inventory-Revised on a sample of federal probationers. *Federal Probation*, *70*(2), 44-78.
- Gendreau, P., Goggin, C. E., & Law, M. A. (1997). Predicting prison misconducts. *Criminal Justice and Behavior*, *24*(4), p414 - 431.
- Gottfredson, S. D., & Moriarty, L. J. (2006). Statistical risk assessment: Old problems and new applications. *CRIME & DELINQUENCY*, *52*, 178-200.
- Grann, M., Belfrage, H., & Tengstrom, A. (2000). Actuarial assessment of risk for violence: Predictive validity of the VRAG and the historical part of the HCR-20. *Criminal Justice and Behavior*, *27*, 97-114.

- Hare, R. D. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, Canada: Multi-Health Systems.
- Harrell, F. E., Jr. (2001). *Regression modeling strategies*. New York: Springer.
- Herrenkohl, T. I., Maguin, E., Hill, K. G., Hawkins, J. D., Abbott, R. D., & Catalano, R. F. (2000). Developmental risk factors for youth violence. *Journal of Adolescent Health, 26*, 176–186.
- Hosmer, D. W., Jr., & Lemeshow, S. (2000). *Applied logistic regression* (2nd ed.). New York: John Wiley.
- Kroner, C., Stadtland, C., Eidt, M., & Nedopil, N. (2007). The validity of the violence risk appraisal guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour and Mental Health, 17*(2), 89 - 100.
- le Cessie, S., & van Houwelingen, J. C. (1992). Ridge estimators in logistic regression. *Applied Statistics, 41*, 191-201.
- Moffitt, T. E. (1993). Adolescence-limited and life-course persistent antisocial behavior: A developmental taxonomy. *Psychological Review, 100*(4), 674–701.
- Nelson, M., Deess, P., & Allen, C. (1999). *The first month out: Post-incarceration experiences in New York City* [Unpublished monograph]. New York.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk*. Washington, DC: American Psychological Association.
- Simourd, D. J. (2004). Use of dynamic risk/need assessment instruments among long-term incarcerated offenders. *Criminal Justice and Behavior, 31*, 306–323.
- Stouthamer-Loeber, M., Loeber, R., Wei, E., Farrington, D. P., & Wikstrom, P. H. (2002). Risk and promotive effects in the explanation of persistent serious delinquency in boys. *Journal of Consulting and Clinical Psychology, 70*(1), 111–123.
- Tolan, P. H., & Gorman-Smith, D. (1998). Development of serious and violent offending careers. In R. Loeber & D. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 68–85). Thousand Oaks, CA: Sage.