EVALUATING THE PREDICTIVE VALIDITY OF THE COMPAS RISK AND NEEDS ASSESSMENT SYSTEM

TIM BRENNAN
WILLIAM DIETERICH
BEATE EHRET
Northpointe Institute for Public Management Inc.

This study examines the statistical validation of a recently developed, fourth-generation (4G) risk–need assessment system (Correctional Offender Management Profiling for Alternative Sanctions; COMPAS) that incorporates a range of theoretically relevant criminogenic factors and key factors emerging from meta-analytic studies of recidivism. COMPAS's automated scoring provides decision support for correctional agencies for placement decisions, offender management, and treatment planning. The article describes the basic features of COMPAS and then examines the predictive validity of the COMPAS risk scales by fitting Cox proportional hazards models to recidivism outcomes in a sample of presentence investigation and probation intake cases ($N = 2,328$). Results indicate that the predictive validities for the COMPAS recidivism risk model, as assessed by the area under the receiver operating characteristic curve (AUC), equal or exceed similar 4G instruments. The AUCs ranged from .66 to .80 for diverse offender subpopulations across three outcome criteria, with a majority of these exceeding .70.

Keywords: COMPAS; predictive validity; survival analysis; risk assessment; probation; area under the curve (AUC); criminogenic needs

In a recent review of the state of the art in correctional assessment, Andrews, Bonta, and Wormith (2006) identified Correctional Offender Management Profiling for Alternative Sanctions (COMPAS; Northpointe Institute for Public Management, 1996) as an example of an emerging fourth-generation (4G) approach to correctional assessment. They also noted that the available 4G approaches, with the exception of the Level of Service/Case Management Inventory (LS/CMI; Andrews, Bonta, & Wormith, 2004), are relatively new and that validation evidence is still required for these newer approaches. This article assesses several key aspects of scale reliability and validity for the COMPAS system.

TRENDS IN CORRECTIONAL ASSESSMENT

The past three decades in correctional practice have seen a progression from first-generation (1G) to currently emerging 4G assessment approaches (Andrews et al., 2006; Blanchette & Brown, 2006; Bonta, 1996; Clements, 1996). These developments occurred as successive generations of assessment and classification methods addressed the more obvious weaknesses of prior phases. These phases and their main characteristics are described below.

The 1G approach relied on clinical and professional judgment in the absence of any explicit or objective scoring rules. It dominated corrections for several decades and remains preferred by many correctional decision makers (Boothby & Clements, 2000; Wormith, 2001). Its weaknesses include excessive subjectivity, inconsistency, bias and potential

stereotyping, legal vulnerability, and lower predictive validity than structured objective methods (Brennan, 1987; Grove & Meehl, 1996; Hastie & Dawes, 2001).

Second-generation (2G) assessments adopted an empirical approach that mainly relied on simple additive point scales, often with only a few standardized factors (e.g., Austin, 1983; S. D. Gottfredson, 1987; Hoffman, 1994). These mostly reflected Dawes's (1979) description of "improper" linear models (p. 571) because the selected factors and weightings were often established by common sense or professional consensus rather than by statistical methods. These methods primarily focused on risk prediction, brevity, and efficiency. The main criticisms included lack of theoretical background, limited coverage of risk and need factors, neglect of dynamic (changeable) risk factors, lack of treatment implications, weak explanatory value, and questionable relevance for female offenders (Blanchette & Brown, 2006; Jones, 1996). However, as noted by Dawes, these linear models are often surprisingly effective in terms of predictive validity and generally outperformed professional judgment or the opinions of trained experts (Grove & Meehl, 1996; Hastie & Dawes, 2001; Mossman, 1994).

Third-generation (3G) assessments of the late 1970s and 1980s introduced a more explicit, empirically based, and theory-guided approach and a broader selection of criminogenic factors. In addition, some of these factors were designed to be dynamically sensitive to change. The Level of Service Inventory–Revised (LSI-R; Andrews & Bonta, 1995) exemplified these trends and perhaps has become the most widely used risk and need assessment in corrections. However, 3G methods, including the LSI-R, eventually were criticized for a narrow theoretical focus (mainly social learning theory), a failure to address gender sensitivity, a dominant focus on risk, and failure to assess offender strengths or protective factors as emphasized in the "good lives" model (Andrews et al., 2006; Blanchette & Brown, 2006; Bloom, 2000; Reisig, Holtfreter, & Morash, 2006; Ward & Stewart, 2003).

Regarding 4G assessments, Andrews et al. (2006) identified several instruments as representing this category, including the Correctional Assessment and Intervention System (National Council on Crime and Delinquency, 2006), LS/CMI, and COMPAS. Several general features appear to characterize 4G approaches. These include (a) a broader selection of explanatory theories, (b) broader range of risk and need factors (content validity), (c) incorporation of the strengths or resiliency perspective, (d) more advanced statistical modeling, (e) seamless integration of the need or risk domain with the agency management information system, and (f) criminal justice databases and Web-based implementation of assessment technology. Such integration allows users to track offenders from intake to case closure to support sequential case management monitoring, information feedback, and decision making. COMPAS has incorporated all of these features; interested readers may obtain full details in Brennan, Dieterich, and Oliver (2007).

The goals of this article are threefold. First, it describes the general design features and technical overview of the COMPAS system. Second, it assesses the reliabilities of the COMPAS scales for both male and female offenders. Third, it assesses the predictive validity of the COMPAS scales for both males and females.

BASIC DESIGN FEATURES OF COMPAS

COMPAS is an automated decision-support software package that integrates risk and needs assessment with several other domains, including sentencing decisions, treatment

and case management, and recidivism outcomes. Documentation of the full software functionality is available at www.northpointeinc.com, and detailed information on the various risk and need scales is provided in the appendix. Beyond the integration of separate databases, the following design features of COMPAS further advance and support evidence-based practice (EBP) in criminal justice agencies.

THEORY-GUIDED ASSESSMENT

Ideally, explanatory theories of criminality should guide the selection of scale content of an assessment system. Criminologists have long lamented the lack of theory-guided assessments (Bonta, 2002; Clements, 1996; Jones, 1996). Thus, 4G systems have a strong emphasis on theory-guided assessment. In contrast to the LSI-R, which was designed primarily around a social learning explanation (Andrews & Bonta, 1998), COMPAS broadens the theoretical coverage to include key constructs from low self-control theory, strain theory or social exclusion, social control theory (bonding), routine activities–opportunity theory, subcultural or social learning theories, and a strengths or good lives perspective.

BROADBAND COMPREHENSIVE ASSESSMENT

Second, 4G approaches introduce a broader comprehensive coverage of criminogenic factors to match the theoretical and explanatory complexity of criminal behavior and to provide sufficient explanatory information to guide case interpretation and intervention planning. Thus, COMPAS includes both theoretically relevant factors and the critical eight criminogenic predictive factors that emerged from recent meta-analytic studies (Andrews et al., 2006; Gendreau, Little, & Goggin, 1996; Lipsey & Derzon, 1998; Lösel, 1995). The 2G approaches generally reflected the opposite trend by minimizing and simplifying assessment to reduce workload burden on staff, which, not unexpectedly, resulted in extreme poverty of explanatory components and an almost total lack of treatment guidance (Austin, 1983; Glaser, 1987; Palmer, 1992).

INTEGRATION OF THE STRENGTH OR RESILIENCY PERSPECTIVE

The strength-based or good lives approach (Andrews et al., 2006; Ward & Brown, 2004) is a natural extension of the shift toward more comprehensive assessment. In their review, Andrews et al. (2006) suggested that measures of strengths and well-being are "highly relevant" (p. 23) for correctional assessments. To address this issue, COMPAS includes a number of strength and protective factors that have shown empirical support for potential risk reduction and protecting offenders from the full impact of criminogenic needs. These include job and educational skills, history of successful employment, adequate finances, safe housing, family bonds, social and emotional support, noncriminal parents and friends, and so on.

MORE ADVANCED STATISTICAL MODELS

4G assessments, in contrast to earlier approaches, are beginning to use more advanced statistical methods for predictive modeling and classification. Although Burgess-type, equally weighted, linear models have performed reasonably well (S. D. Gottfredson, 1987; Mossman, 1994), powerful multivariate, model-averaging, mixed-model ensemble methods

and artificial intelligence are now entering correctional assessment approaches. For example, the COMPAS risk and classification models use logistic regression, survival analysis, and bootstrap classification methods in a broad repertoire of prediction and classification procedures (Brennan, Breitenbach, & Dieterich, 2008). The present article specifically examines its predictive models for recidivism based on survival analysis (Cox regression).

INTEGRATION WITH CRIMINAL JUSTICE DATABASES TO FACILITATE EBP

Another feature of 4G methods, including COMPAS, is the seamless integration of the risk and needs domain with separate domains of sentencing decisions, institutional processing and placement decisions, case management decisions, treatments given (type and amount), and various outcomes (across time). This integration provides support for correctional agencies in implementing EBP studies (see Andrews et al., 2006; Brennan, Wells, & Alexander, 2004).

The COMPAS system includes two additional design features of some note: a treatment-explanatory classification to support staff with specific responsivity decisions and gender-sensitive calibration, described below.

TREATMENT-EXPLANATORY CLASSIFICATION TO ADDRESS SPECIFIC RESPONSIVITY

Andrews et al. (2006) suggested that specific responsivity of offenders is the least explored of their risk–need–responsivity principles. Yet specific responsivity is a critical and recurrent challenge for treatment providers in matching individual offenders to appropriate treatment regimes (Brennan, 2008a; Meier, 2002; Millon & Davis, 1997; Warren & Hindelang, 1979). COMPAS addresses specific responsivity and client–treatment matching using two well-known approaches. First, it provides a person-centered assessment chart of decile scores for each risk and need scale. Second, following the lineage of Marguerite Warren, Ted Palmer, and others (also see Harris & Jones, 1999; Van Voorhis, 1994), COMPAS provides a treatment-relevant typology that integrates risk and need. This explanatory typology identifies and demarcates several specific pathways that may guide differential targeting and programming for diverse offender types who belong in one particular pathway. Although the present article does not address this treatment-relevant taxonomy, detailed descriptions of these pattern-seeking methods are presented in Brennan et al. (2008) and Brennan (2008b).

GENDER-SENSITIVE ASSESSMENT

A major criticism of 2G and 3G approaches is that they largely based their assessment and classification methods on dominantly male samples and then mechanically applied these to female offenders (Blanchette & Brown, 2006; Bloom, 2000; Brennan, 2008b; Farr, 2000; Hannah-Moffatt & Shaw, 2001). However, compelling arguments are now advanced for systematic validation of instruments on separate female and male offender samples (Hardyman & Van Voorhis, 2004; Holtfreter & Cupp, 2007). COMPAS addresses this issue, first by using separate samples of males and females to develop gender-specific calibrations of all risk and need factors and second by evaluating its predictive and classification models on separate male and female samples (see below). Plans are also under way to incorporate additional gender-specific factors from recent research on gender-sensitive risk and need factors into COMPAS (Salisbury, Van Voorhis, & Wright, 2006).

METHOD

PARTICIPANTS

Participants were 2,328 individuals who were assessed with COMPAS as part of their processing at entry into probation agencies. All individuals with complete data who were administered the full COMPAS assessment were included. The sample represented about 15% of all COMPAS assessments conducted at these agencies during this period. Individuals were excluded if they were missing data (25%) or if they were not administered the full COMPAS (60%). Women composed 19% of the sample. The ethnic composition of the sample is 76% White, 15% African American, 7% Latino, and 2% Other. The average age of participants was 31.9 years (range = 18.0 to 69.7). In the sample overall, 45% of the presenting offenses were misdemeanors, 48% nonassaultive felonies, and 7% assaultive felonies. The median number of prior arrests was 3 (range = 0 to 57). Among the probation cases, 9% were split-sentence cases (jail and probation).

ADMINISTRATION

The assessments were conducted by local probation officers between January 2001 and December 2004 at 18 county-level probation agencies in an eastern state. Interviews were conducted at the point of presentence investigation (PSI) or at probation intake (approximately 50% each). Staff and supervisors take a 2-day COMPAS training program that covers relevant interview techniques, response categories, item meanings, and quality assurance issues. Official criminal records are used to complete the current offense and criminal history sections of COMPAS prior to the interviews. The interviews typically require approximately 45 to 60 min, depending on the extent of probing.

MEASURES

COMPAS scales. This study examined the predictive validity of all the COMPAS base scales (listed in Table 1) and also the main Recidivism Risk Scale. The Recidivism Risk Scale is a regression model that has been used in COMPAS since 2000. This regression model was trained to predict new offenses in a probation sample. The system transforms a linear predictor from the regression model to a decile score. The system calculates a recidivism risk decile score by referencing an appropriate COMPAS norm group. For the current analyses, a gender-specific composite norm group was used. The composite norm group ($n = 7,381$) was constructed from COMPAS assessment data collected in prison (34%), jail (14%), and probation (53%). The set of base scales included criminal involvement, history of noncompliance, history of violence, current violence, criminal associates, substance abuse, financial problems, vocational or educational problems, family criminality, social environment, leisure, residential instability, social isolation, criminal attitudes, and criminal personality.

Dependent variables: Three outcomes for survival analyses. We matched COMPAS assessment data with computerized official criminal history records and constructed multiple-record survival data sets using assessment and event dates in the criminal history data. These included crime dates, arrest dates, dispositions, disposition dates, sentence type, and sentence length. The three outcomes selected as dependent variables for this study included

TABLE 1:   Alpha Coefficients and Their Differences Between Women and Men, With Pointwise 95% Confidence Intervals (CIs)

	α	Women	Men	Difference	Lower Bound 95% CI	Upper Bound 95% CI
Criminal Involvement	.87	.85	.87	−.02	−.04	.01
History of Noncompliance	.68	.62	.67	−.04	−.11	.02
History of Violence	.73	.70	.73	−.03	−.07	.02
Current Violence	.59	.62	.59	.03	−.03	.09
Criminal Associates	.68	.68	.68	.00	−.05	.05
Substance Abuse	.79	.81	.78	.03	.00	.06
Financial Problems	.73	.75	.72	.03	−.02	.07
Vocational or Educational	.71	.73	.71	.02	−.02	.06
Family Criminality	.63	.64	.63	.02	−.04	.07
Social Environment	.74	.70	.74	−.04	−.09	.01
Leisure	.82	.81	.83	−.02	−.05	.01
Residential Instability	.65	.61	.65	−.05	−.11	.01
Social Isolation	.81	.82	.81	.01	−.02	.04
Criminal Attitudes	.82	.82	.82	.01	−.02	.03
Criminal Personality	.76	.76	.76	.00	−.03	.04

Note. The difference in alphas is significant if the 95% confidence interval does not include zero.

(a) an arrest for any offense, (b) an arrest for a person offense, and (c) an arrest for a felony offense.

We defined an offense as a finger-printable arrest involving a charge and filing for any uniform crime reporting (UCR) code. A person offense is a finger-printable arrest involving a charge and filing for any UCR code for murder, voluntary manslaughter, forcible rape, robbery, aggravated assault, simple assault, burglary (with weapon of an occupied dwelling), dangerous weapons, sex offenses, extortion, arson, and kidnap. This category includes misdemeanor and felony offenses.

ANALYSIS

We fitted separate cause-specific Cox proportional hazard models to each recidivism outcome (Kalbfleisch & Prentice, 2002). Analysis time is the number of days from COM-PAS assessment date to first failure or end of study, whichever occurred first. As mentioned, the assessments were conducted between January 2001 and December 2004. The end of study is the date of the recidivism outcomes computer match (March 3, 2006). We determined the failure time point from the offense date associated with the recidivism outcome of interest. Cases remained in the risk set and contributed information to the analyses until the point of failure or end of study, whichever occurred first. The models controlled for intermittent periods of incarceration during the follow-up by removing the case from the risk set during these intermittent gaps. We also removed split-sentence cases from the risk set during the time they were in jail. We removed PSI cases that received a subsequent jail or prison sentence from the risk set during the incarceration period. In the felony offense model, there were 433 cases with a gap, with an average time on gap of 206 days (range = 1 to 1,440 days). The median time at risk in the felony offense model was 759 days (range = 1 to 1,722 days).

First, we fitted a series of univariable Cox survival models in which the hazard for each of the three recidivism outcomes was regressed on each COMPAS base scale. These

models were fitted in three partitions of the sample: the full sample, men only, and women only. Next, two multivariate models were fitted to each recidivism outcome in each partition. Model I included all the COMPAS base scales. Model II included all the COMPAS base scales plus age at first arrest. Finally, a model that included only the Recidivism Risk Scale was fitted to each recidivism outcome in each partition.

To gauge the predictive utility of the above two COMPAS multivariate base scale models and the Recidivism Risk Scale model, we estimated the area under the receiver operating characteristic curve (AUC) for all three models. For survival models, the most relevant measure of predictive discrimination is the concordance index, which is equivalent to the AUC and is defined as the probability that the predictor values and survival times for a pair of randomly selected cases are concordant. A pair is concordant if the case with the higher predictor value has a shorter survival time. The calculation is based on the number of all possible pairs of nonmissing observations for which survival time can be ordered and the proportion of relevant pairs for which the predictor and survival time are concordant (Harrell, Califf, Pryor, Lee, & Rosati, 1982).

RESULTS

RELIABILITY

Table 1 provides reliability coefficients (Cronbach's α) to indicate the internal consistency of the core COMPAS scales both for the total sample and by gender. Alpha is the most widely used measure of internal consistency of summative scales. By convention, alpha coefficients of .70 or higher indicate satisfactory reliability. The table indicates that a large majority of these alphas are in the satisfactory range of close to or above .70 with only a few exceptions (current violence, family criminality, and residential instability), and these are close to an acceptable range. The alpha coefficients were statistically equivalent for both genders.

PREDICTIVE VALIDITY

Table 2 provides a summary of the survival experience of men and women for the three outcomes examined in the study. For each offense type, the table shows the number of failures that occurred during each year and the estimated survivor function at the end of each year. The survivor function is the probability of surviving beyond Time *t,* given survival up to Time *t.* The survivor function is interpreted as the cumulative proportion surviving over time. Note that only the first 4 years of the follow-up are shown.

Table 3 shows the results from univariable Cox regressions of the hazard for a new felony offense on each COMPAS base scale. The results indicate that all except three of these base scales reach significant levels in predicting felony recidivism. The scales that did not attain significance were current violence, financial problems, and residential instability. Note that substance abuse has a negative effect on felony recidivism. This may have resulted from the selected outcome variable (felony recidivism) for this analysis. Drug offenders (overall) may be at lower risk for new arrests for serious felony offenses.

Table 4 shows the results of the multivariate Cox regression of the hazard of a felony offense on the COMPAS base scales. In this model, which combines all COMPAS base

TABLE 2: Number Failing Each Year and Survivor Function Through the End of Each Year for Any Offense, Offenses Against Persons, and Felony Offenses for Men and Women

	Women (n = 449)		Men (n = 1,879)	
Offense Type	Number Failing Each Year	Survivor Function	Number Failing Each Year	Survivor Function
Any				
1st year	64	.85	308	.81
2nd year	32	.76	180	.69
3rd year	14	.70	63	.62
4th year	10	.58	41	.52
Person				
1st year	11	.97	114	.93
2nd year	10	.95	90	.87
3rd year	5	.92	33	.83
4th year	2	.89	23	.78
Felony				
1st year	16	.96	134	.92
2nd year	12	.93	95	.85
3rd year	7	.89	31	.82
4th year	1	.88	20	.77

Note. The description of the survival experience is limited to the first 4 years of the follow-up.

TABLE 3: Results From Univariable Cox Proportional Hazards Models Regressing the Hazard for a Felony Offense on Each COMPAS Base Scale

	coeff	exp(coeff)	SE(coeff)	p Value
Criminal involvement	0.033	1.03	0.013	.008
History of noncompliance	0.148	1.16	0.020	.000
History of violence	0.108	1.11	0.018	.000
Current violence	0.101	1.11	0.052	.052
Criminal associates	0.148	1.16	0.022	.000
Substance abuse	−0.091	0.91	0.023	.000
Financial problems	0.009	1.01	0.024	.691
Vocational or educational	0.118	1.13	0.014	.000
Criminal attitudes	0.057	1.06	0.009	.000
Family criminality	0.100	1.11	0.036	.006
Social environment	0.257	1.29	0.032	.000
Leisure	0.073	1.08	0.016	.000
Residential instability	0.016	1.02	0.014	.277
Criminal personality	0.048	1.05	0.007	.000
Social isolation	0.029	1.03	0.010	.003

scales, the significant predictors of felony recidivism are history of noncompliance, criminal associates, substance abuse, financial problems, vocational or educational, and high crime (social) environment. As often occurs when multiple predictor variables are used and when some are correlated, the signs of certain parameters can take unexpected directions.

Table 5 shows the results of the Cox regression of the hazard for a new felony offense on the levels of the Recidivism Risk Scale. The hazard of a new felony offense for cases that score high on the Recidivism Risk Scale is 5.66 times the hazard of cases that score low (p value < .001). The hazard of the high-risk group relative to the medium-risk group is 1.84 (p value < .001).

TABLE 4: Results From Multivariable Cox Proportional Hazards Model Regressing the Hazard for a Felony Offense on the COMPAS Base Scales

	coeff	exp(coeff)	SE(coeff)	p Value
Criminal involvement	0.003	1.00	0.018	.854
History of noncompliance	0.124	1.13	0.032	.000
History of violence	0.026	1.03	0.024	.276
Current violence	0.001	1.00	0.057	.982
Criminal associates	0.072	1.07	0.028	.009
Substance abuse	−0.134	0.87	0.027	.000
Financial problems	−0.059	0.94	0.026	.025
Vocational or educational	0.081	1.08	0.017	.000
Criminal attitudes	0.014	1.01	0.011	.212
Family criminality	−0.001	1.00	0.040	.978
Social environment	0.117	1.12	0.037	.002
Leisure	0.015	1.01	0.021	.476
Residential instability	−0.002	1.00	0.015	.880
Criminal personality	0.007	1.01	0.011	.555
Social isolation	−0.003	1.00	0.012	.771

TABLE 5: Results From Cox Proportional Hazards Model Regressing the Hazard for a Felony Offense on the Recidivism Risk Scale

	coeff	exp(coeff)	SE(coeff)	p Value	95% Confidence Interval
Medium risk	1.12	3.07	0.473	<.001	2.27, 4.15
High risk	1.73	5.66	0.843	<.001	4.22, 7.58

Note. The reference category is the low risk level of the Recidivism Risk Scale.

Figure 1 shows a plot of the Nelson–Aalen (Aalen, 1978; Nelson, 1972) estimator of the cumulative hazard function $\hat{H}(t)$ within levels of the Recidivism Risk Scale. The estimator is defined as

$$\hat{H}(t) = \sum_{j:t_j \leq t} \frac{d_j}{n_j}, \tag{1}$$

where n_j stands for the number of cases in the risk set just before Time t_j and d_j is the number of failures at Time t_j. The cumulative hazard is the expected number of failures for an individual as a function of time, if failure was a repeatable process. Figure 1 also shows the size of the risk set (n_j) at 180-day intervals in each level of the Recidivism Risk Scale. Although the maximum follow-up time in the data is 1,887 days, the time axis in the plot is truncated to 1,620 days because the risk set is small and the estimates less precise beyond this point.

Finally, we assessed the discriminatory power of the three models for predicting the three criteria of interest (general offenses, offenses against persons, and felony offenses) in each of the three partitions of the sample. As described previously, Model I includes all COMPAS base scales, Model II adds age at first arrest to these COMPAS base scales, and Model III represents the COMPAS Recidivism Risk Scale. Again, because we are fitting survival models, we estimate Harrell's concordance index (Harrell et al., 1982) and interpret it as the AUC. A rule of thumb according to several recent articles is that AUCs of .70
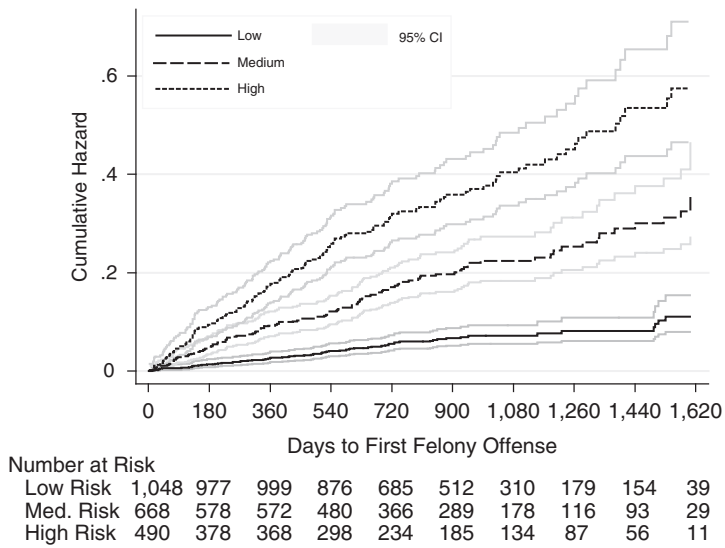
Figure 1:   Nelson-Aalen Plot of the Cumulative Hazard for a Felony Offense in Each Level of the Recidivism Risk Scale, With Pointwise 95% Confidence Bands

Note. CI = confidence interval.

TABLE 6:   Area Under the Curve Values for COMPAS Risk Models Predicting Any Offense, Offenses Against Persons, and Felony Offenses

Model	Total Sample (N = 2,328)			Women (n = 449)			Men (n = 1,879)		
	Any	Person	Felony	Any	Person	Felony	Any	Person	Felony
COMPAS I	.66	.72	.70	.69	.78	.68	.67	.71	.71
COMPAS II	.68	.73	.72	.72	.80	.69	.68	.72	.73
Recidivism Risk III	.68	.71	.70	.65	.76	.66	.68	.70	.71

Note. COMPAS Model I includes COMPAS base scales; COMPAS II adds age at first arrest to Model I.

or above typically indicate satisfactory predictive accuracy, and measures between .60 and .70 suggest low to moderate predictive accuracy (Aos & Barnoski, 2003; Jones, 1996; Quinsey, Harris, Rice, & Cormier, 1998).

Table 6 presents the model-specific AUCs for all of these analyses. The AUC values range from .66 to .80, with a majority being above .70, which suggests satisfactory predictive validities of these COMPAS risk models for all three recidivism outcomes. The AUCs for person and felony offenses are a little higher than those for the broader and less precisely defined recidivism category "any offense" (which includes both misdemeanor and felony offenses). The AUCs for predicting person offenses range from .71 to .80, with high AUC values for women (.76 to .80). However, for women, although the sample size was 449, there were only 29 women with a person offense, which reduces statistical power and suggests caution regarding this result.

TABLE 7: Area Under the Curve Values for COMPAS Risk Models Predicting Any Offense, Offenses Against Persons, and Felony Offenses for White and African American Men

Model	White Men (n = 1,412)			African American Men (n = 296)		
	Any	*Person*	*Felony*	*Any*	*Person*	*Felony*
COMPAS I	.69	.74	.73	.64	.69	.69
COMPAS II	.71	.75	.75	.66	.71	.72
Recidivism Risk III	.69	.71	.71	.67	.72	.73

Last, we examined the predictive accuracy of each of the models for African American men and White men. We do not report results for African American women because the effective sample sizes for most of the outcomes were too small to calculate separate AUCs for that group. Table 7 presents the results from each model for the outcomes any arrest, person offense arrest, and felony arrest for African American and White men. The AUCs for African American men range from .64 to .73. As was the case in the full sample, the highest AUCs are obtained for the felony offense and person offense arrest outcomes. The AUC results for White men are quite similar to the results for African American men, except that White men have somewhat higher AUCs on the COMPAS base scale models.

DISCUSSION AND CONCLUSIONS

This study examined the reliability (internal consistency) and predictive validity of the COMPAS risk and needs scales on a large sample of PSI and probation cases. A first general conclusion is that a majority of these scales reach levels of internal consistency and predictive validity that are within generally acceptable ranges. Second, the separate uni-variable analyses show that a majority of the specific COMPAS risk and need base scales were significantly associated with felony recidivism (Table 3). Third, results from survival analyses demonstrated that the predictive power of the three models tested was comparable to, and in some cases higher than, similar risk predictive instruments in this field. We now examine some more detailed specific findings.

Regarding internal consistency, most of the scales have alpha coefficients equal to or greater than .70, with only three exceptions, and the latter were close to acceptable levels. These satisfactory results reemerged in both male and female subsamples. We found no significant differences in alpha levels between male and female subsamples, suggesting that the scales are equally reliable for men and women.

Regarding predictive validity, we must first place the present results in the context of recent research and accuracy performances for offender recidivism studies. An important point is that the AUC has become the preferred measure of accuracy largely because of its independence across base rates and selection ratios that allow it to provide clearer comparisons across different predictive instruments and studies (Flores, Lowenkamp, Smith, & Latessa, 2006; Quinsey et al., 1998). Another contextual issue concerns interpretations of levels of the AUC coefficient. As noted previously, AUCs in the .50s are considered to have little or no predictive accuracy, those in the .60s are considered weak, those approaching or above the .70s are moderate, and those in the .80s are strong (Tape, 2003). However, various authors appear to use

different standards. For example, Flores et al. (2006) described their achieved AUC of .689 for the LSI-R as valid and robust and as moderate to large. Similarly, Kroner, Stadtland, Eidt, and Nedopil (2007) describe an AUC of .703 as representing a high predictive accuracy in a study of the Violence Risk Appraisal Guide (VRAG). Perhaps the relative recency of using the AUC to investigate offender risk assessment explains variations in evaluative statements. Finally, with more studies reporting AUCs in this area, it appears that the accuracy levels achieved for most current instruments, across a variety of samples and outcome variables, generally fall in the range of .65 to .75, with only a few exceptions (see below).

The present study of COMPAS models produced AUCs mostly in a range of .70 to .80. Specifically, 16 out of 27 cells examined for AUC reached .70 or above, with a smaller set of cells in the .66 to .69 range. Note in particular the predictive accuracies of COMPAS for person offenses in which all 9 of the cells (total sample, males or females, and three models) had AUCs between .70 and .80. Thus, we may conclude that COMPAS predictive accuracies are similar to or slightly higher than AUCs obtained by other major instruments in this field (e.g., Barbaree, Seto, Langton, & Peacock, 2001; Barnoski & Aos, 2003; Dahle, 2006; Flores et al., 2006; Grann, Belfrage, & Tengstrom, 2000; Quinsey et al., 1998). Furthermore, the present findings are in the same range as found in the initial validation studies of the COMPAS recidivism risk model for probationers that produced AUCs of .72 and .74 over a 24-month outcome period (Brennan, Dieterich, & Oliver, 2004).

The AUCs of the other main instruments often used for offender risk prediction may further help to contextualize the above findings. Perhaps the best known instruments are the VRAG (Quinsey et al., 1998), the 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, as illustrated below.

VRAG

Quinsey et al. (1998) found an AUC of .76 in a large-scale, multiyear recidivism study. Barbaree et al. (2001) reported AUCs of .69 in predicting serious reoffending and .77 when predicting any reoffense for sex offenders. Kroner et al. (2007) obtained an AUC of .703 in a study of reoffending among mentally ill offenders.

LSI-R

The recent review by Andrews et al. (2006) did not provide AUCs for the LSI-R. However, Barnoski and Aos (2003) found AUCs of .64 to .66 for the LSI-R in predicting felony and violent recidivism among Washington State prisoners. Flores et al. (2006) found an AUC of .689 using the LSI-R to predict reincarceration among federal probationers. Dahle (2006) reported an AUC of .65 using the LSI-R to predict violent recidivism. Barnoski (2006) reported an AUC of .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 et al. (2000) found AUC levels of .64 to .75 based on various follow-up time frames. Barbaree et al. (2001) reported AUCs of .61, .65, and .71 for the PCL-R in predicting various recidivism outcomes among sex offenders.

The above findings clearly do not exhaust the full range of studies in this area. As more studies report AUCs for specific instruments, varying populations, outcome variables, and time frames, it may become possible to identify which instruments perform well in these varying conditions.

The present study has several strengths and limitations. Strengths include the large sample ($N = 2,328$), a multiyear outcome period, and an examination of AUCs for different COMPAS models, different offense categories, and across genders. The incorporation of survival modeling also shows not only that these AUCs achieved significant discrimination between recidivists and nonrecidivists but also that the timing of failure events can be predicted using either the COMPAS base scales model or the overall recidivism risk model. However, the present study did not systematically address variations in predictive accuracy by offender subgroups broken down by age, ethnicity, and race; level of addiction; length of follow-up; and so on. Several large-scale studies of COMPAS that will allow such detailed examination are in progress. The following issues may still require further research.

PREDICTIVE ACCURACY BY OUTCOME OFFENSE

The present study found minor differences in AUCs for COMPAS risk models in predicting differing outcomes and clearly achieved stronger results for new offenses against persons and for felony offenses than for the broader category of any new offense. This may stem from the higher precision of definitions for the first two outcomes as opposed to the less precise "any" new offense, as this included both misdemeanors and felonies across a wide range of offenses. Barnoski and Drake (2007) found similar results when they examined the validity of a static risk scale for predicting three outcome categories (violent, property and violent, and felony).

PREDICTIVE ACCURACY BY GENDER

The present study found that the basic COMPAS recidivism model predicts behaviors for men and women about equally well, again similar to Barnoski and Drake (2007), who found only small differences by gender. We realize that there are substantial concerns in this topic, and a search for gender-sensitive risk factors is currently under way. COMPAS is being augmented by additional risk and need factors of specific importance for female offenders (Blanchette & Brown, 2006; Brennan, 2008b; Salisbury et al., 2006).

PREDICTIVE ACCURACY BY ETHNICITY

The present study found that the COMPAS recidivism models preformed equally well for African American and White men at predicting the arrest outcomes. There is only one previous study of which we are aware that examined the predictive accuracy of the COMPAS for different ethnic groups, and that study reported much weaker results for African American men (Fass, Heilbrun, DeMatteo, & Fretz, 2008). In predicting rearrest within 1 year of release, Fass et al. (2008) reported AUCs for the COMPAS Recidivism Risk Scale of .81 for Whites, .67 for Hispanics, .48 for African Americans, and .53 for the total sample assessed with COMPAS ($N = 276$). However, their study has at least one major weakness that renders its findings unreliable. Their small overall sample size and base rates resulted in extremely small effective sample sizes for the ethnic groups (African American = 36, Hispanic = 4, White = 1), and this almost ensures unreliable results.

PREDICTIVE ACCURACY ACROSS DIVERSE CRIMINAL JUSTICE POPULATIONS

The present study did not include other specific criminal justice populations from prison, parole, community corrections, or jails. However, a current, large-scale, parole reentry study is evaluating predictive performance for parolees released to the community (Zhang, Farabee, & Roberts, 2007). Preliminary findings reported AUCs of .67 rising to .72 with only minor adjustments to the COMPAS recidivism risk model.

In conclusion, given that instrument validation is an ongoing process, we acknowledge that numerous further tests and models could be applied to examine the predictive validity of COMPAS risk models. The present results, however, are encouraging and suggest that the COMPAS risk models reach levels of reliability, predictive validity, and generalizability that are at least equal to those of other major instruments in offender risk assessment.

APPENDIX
Scale Content and Selection: Theoretical and Empirical Background

This appendix describes the item content and research background for each Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) scale listed in Table 1. Full details of psychometric properties, theoretical justification, and supportive empirical studies for each scale are given in Brennan, Dieterich, and Oliver (2007). We now describe the background of the COMPAS scales, along with the main items loading on each scale and factor loadings (in parentheses).

Criminal Involvement. This scale includes items pertaining to number of prior arrests and convictions, frequency of incarceration, and criminal justice involvements. The highest loading item is total number of prior arrests (.52). Past criminal involvement has been consistently supported by meta-analytic studies as a major risk factor for predicting ongoing criminal behavior (Andrews & Bonta, 1998; Gendreau, Little, & Goggin, 1996).

History of Violence. This scale includes official history items reflecting prior arrests and convictions for violent felonies, use of weapons, infractions for fighting, and so on. The highest loading items are the number of prior assaultive felony convictions (.47) and frequency of injury to victims (.40). The research literature has indicated that the likelihood of future violence appears to increase with each instance of a prior violent incident (Farrington, 1991; Lipsey & Derzon, 1998; Parker & Asher, 1987).

History of Noncompliance. This scale includes official items reflecting failures to appear, failures of drug tests, failures to comply with sentencing conditions, revocations for technical reasons, and so forth. High-loading items include the number of probation revocations (.56) and prior failures to appear (.37). Repeated noncompliance with criminal justice conditions and treatment regimes has emerged as a predictor of both violent and general recidivism (Stalans, Yarnold, Seng, Olsen, & Repp, 2004).

Criminal Associates. This scale assesses associations with others who are involved in drugs, criminal activity, and gangs. High-loading items include friends who have been arrested (.41) and friends who have been gang members (.43). This construct is of central theoretical importance in both social learning and subcultural theories of crime (Cullen & Agnew, 2003; Elliott, Huizinga, & Ageton, 1985). Meta-analytic research has consistently shown that having antisocial associates is a major risk factor for recidivism (Gendreau et al., 1996).

APPENDIX (continued)

Substance Abuse. The central items in this scale reflect the influence of alcohol or drugs on the current offense (.40), perceived benefit of substance abuse treatment (.41), and prior substance abuse treatment (.37). Drug use has consistently emerged as a significant risk factor for general criminal and violent behavior (National Institute of Justice, 1999; National Research Council, 1993) and is a major risk factor in meta-analytic studies (Gendreau et al., 1996).

Financial Problems and Poverty. This scale includes items such as worry about financial survival (.53), problems paying bills (.52), and not enough money to get by (.52). Although poverty has shown only modest predictive power in meta-analytic studies (Gendreau et al., 1996), decades of research has shown reliable associations between poverty and high crime rates and related risk factors such as unstable residence, family disruption, single-parent families, community disorganization, and substandard housing (National Research Council, 1993; Sampson & Lauritsen, 1994). Theoretically, poverty is a key factor in strain or social marginalization and subcultural theories (Cullen & Agnew, 2003).

Occupational and Educational Resources or Human Capital. This scale includes items reflecting levels of educational and vocational occupational success such as job skills (.43), current unemployment (.52), low wages (.49), and employment history (.39). A social achievement (human capital) scale was selected for both empirical and theoretical reasons (Coleman, 1990; Gendreau et al., 1996; Hagen, 1998). It is central to strain theory because people with lower social capital have fewer life chances and more restricted opportunities than do those with greater capital. The scale is dynamic because human capital can be built or destroyed. Job loss or high school dropout may lower economic and social opportunities, whereas completing job skills training or obtaining a GED may increase these chances.

Family Crime. The COMPAS scale of family criminality includes items assessing the criminality and drug use of mother, father, and siblings. The highest loading items include parent ever jailed (.55), parent has had drug problems (.43), and mother ever arrested (.40). Many empirical studies have linked delinquency and adult crime to antisocial families (Farrington, Jolliffe, Loeber, Stouthamer-Loeber, & Kalb, 2001; Farrington & West, 1993; Lykken, 1995). Social learning theory has also linked deviant behavior to violent or criminal family role models and ineffective parenting (Lykken, 1995). Heritability theories have included gene-based evolutionary theories (Ellis & Walsh, 1997) and biosocial theories of sociopathy (Mealey, 1995).

High Crime Neighborhood. This scale assesses levels of crime (.44), gang activity (.40), and drug activity (.39) in the person's neighborhood. Living in a high-crime neighborhood is an established correlate of both delinquency and adult crime (Sampson & Lauritsen, 1994; Thornberry, Huizinga, & Loeber, 1995). It plays a role in social disorganization, social learning, and subcultural theories (Cullen & Agnew, 2003; Sampson, Raudenbush, & Earls, 1997).

Boredom and Lack of Constructive Leisure Activities (Aimlessness). This scale includes items from two closely linked themes: boredom proneness and lack of engaging leisure activities. Dominant items include often bored (.46), nothing to do (.47), restless with current activities (.47), and scattered attention (.37). Although conceptually different, items from these two themes all load on a single factor. Theoretically, an absence of constructive leisure activities partially reflects weak engagement bonds of early social control theory (Hirschi, 1969), and a similar concept (idle hands) enters routine activities theory (Osgood et al., 1996). Finally, restlessness, distractibility, and attention problems enter M. R. Gottfredson and Hirschi's (1990) low self-control theory and Hare's (1991) related construct of psychopathy.

APPENDIX (continued)

Residential Instability. The present scale includes items assessing the number of recent moves (.39), homelessness (.33), and absence of a verifiable address (.34). The background literature indicates that transience is often associated with poverty, poor housing, social disorganization, and crime (McNaughton, 2007; National Research Council, 1993). Theoretically, transience and homelessness may weaken social ties and have been associated with family breakup, social exclusion, and stressful life events (Marris, 1987). Theoretically, it plays a role in both social control theory (weakening or attenuating social bonds) and strain theory (poverty, personal stress, marginalization). We note that personal stress or distress has emerged as a risk factor with modest predictive validity for recidivism (Gendreau et al., 1996).

Social Isolation Versus Social Support. This scale captures social isolation at one pole and social supports at the other. It includes items indicating self-reported loneliness (.33), absence of friends (.40), feeling left out of things (.33), and no close or best friend (.37). Social support theory suggests that even in high-risk environments social supports may mediate or buffer the criminogenic effects of economic and social strain (Bennett & Morabito, 2006; Estroff, Zimmer, Lachicotte, & Benoit, 1994; National Research Council, 1993; Stevenson, 1998). In addition, at prison release and reentry to society, prisoners with stronger social and family supports are found to have lower recidivism (Solomon, Johnson, Travis, & McBride, 2004). Theoretically, this factor enters both strain theory (buffering strain) and social control theory (reflecting social bonds).

Criminal Attitude. This scale assesses antisocial attitudes using items that may justify, excuse, or minimize damage caused by the offender's crime. Prominent items include the law does not help average people (.32), minor offenses such as drug use don't hurt anyone (.24), and things stolen from rich people won't be missed (.36). Antisocial attitudes have emerged in meta-analytic studies as a major risk factor (Andrews, Bonta, & Wormith, 2006; Gendreau et al., 1996). There is less agreement on the particular attitudes that are most useful or predictive (e.g., minimizing the damage of offenses, tolerance for law violation, etc.; Walters, 1995). In the absence of consensus, COMPAS uses a higher order scale with items adapted from Bandura, Barbaranelli, Caprara, and Pastorelli (1996).

Antisocial Personality. This scale addresses impulsivity, absence of guilt, selfish narcissism, dominance, risk taking, and anger or hostility. Representative items include short temper (.39), often does things without thinking (.30), and seen as cold and callous (.32). It is not designed as a comprehensive coverage of all personality subfactors but as a short, broadband scale using only the first principle component from a larger battery of antisocial personality items adapted from Eysenck and Eysenck (1978) and Bandura et al. (1996). Important subfactors are impulsivity, risk taking, restlessness or boredom, absence of guilt (callousness), selfish narcissism, interpersonal dominance, and anger or hostility (Bandura et al., 1996; Cooke, Forth, & Hare, 1996; Lilienfeld & Andrews, 1996; Marcus, 2003; Widiger & Lynum, 1998). Empirical support for antisocial personality (variously measured) is found in Gendreau et al. (1996), Bandura et al. (1996), Blackburn and Coid (1998), and Quinsey, Harris, Rice, and Cormier (1998). Theoretically, antisocial personality plays an important role in theories of antisocial dispositions (Farrington, 2003; M. R. Gottfredson & Hirschi, 1990; Hare, 1991).

REFERENCES

Aalen, O. O. (1978). Nonparametric inference for a family of counting processes. *Annals of Statistics, 6,* 701-726.
Andrews, D. A., & Bonta, J. (1995). *The Level of Service Inventory–Revised.* Toronto: Multi-Health Systems.
Andrews, D. A., & Bonta, J. (1998). *The psychology of criminal conduct* (2nd ed.). Cincinnati, OH: Anderson.

Andrews, D. A., Bonta, J., & Wormith, J. S. (2004). *The Level of Service/Case Management Inventory (LS/CMI).* Toronto: 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 & Delinquency*, *52*, 7-27.

Aos, S., & Barnoski, R. (2003). *Washington's offender accountability act: An analysis of the Department of Correction's risk assessment* (Document No. 03-12-1202). Washington, DC: Washington Institute for Public Policy.

Austin, J. (1983). Assessing the new generation of prison classification models. *Crime & Delinquency, 29,* 561-576.

Bandura, A., Barbaranelli, C., Caprara, G. V., & Pastorelli, C. (1996). Mechanisms of moral disengagement in the exercise of moral agency. *Journal of Personality and Social Psychology, 71,* 364-374.

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,* 490-521.

Barnoski, R. (2006). *Sex offender sentencing in Washington State: Predicting recidivism based on the LSI-R.* Retrieved January 20, 2008 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). Olympia: Washington State Institute for Public Policy.

Barnoski, R., & Drake, E. K. (2007). *Washington's Offender Accountability Act: Department of Corrections' static risk assessment.* Olympia: Washington State Institute for Public Policy.

Bennett, R. R., & Morabito, M. S. (2005, November). *Institutional social support and crime: A cross-national investigation.* Paper presented at the annual meeting of the American Society of Criminology, Toronto, Canada.

Blackburn, R., & Coid, J. W. (1998). Psychopathy and the dimensions of personality in violent offenders. *Personality and Individual Differences, 25,* 129-145.

Blanchette, K., & Brown, S. L. (2006). *The assessment and treatment of women offenders: An integrative perspective*. New York: John Wiley.

Bloom, B. (2000). Beyond recidivism: Perspectives on evaluation of programs for female offenders in community corrections. In M. McMahon (Ed.), *Assessment to assistance: Programs for women in community corrections* (pp. 107-138). Latham, MD: American Correctional Association.

Bonta, J. (1996). Risk-needs assessment and treatment. In A. T. Harland (Ed.), *Choosing correctional options that work: Defining the demand and evaluating the supply* (pp. 18-32). Thousand Oaks, CA: Sage.

Bonta, J. (2002). Risk needs assessment: Guidelines for selection and use. *Criminal Justice and Behavior, 29,* 355-379.

Boothby, J. L., & Clements, C. B. (2000). A national survey of correctional psychologists. *Criminal Justice and Behavior, 27,* 715-731.

Brennan, T. (1987). Classification: An overview of selected methodological issues. In D. M. Gottfredson & M. Tonry (Eds.), Prediction and classification: Criminal justice decision making (pp. 201-248). Chicago: University of Chicago Press.

Brennan, T. (2008a). Explanatory diversity among female delinquents: Examining taxonomic heterogeneity. In R. Zaplin (Ed.), *Female crime and delinquency: Critical perspectives and effective interventions* (pp. 197-232). Boston: Jones and Bartlett.

Brennan, T. (2008b). Institutional assessment and classification of female offenders: From robust beauty to person-centered assessment. In R. Zaplin (Ed.), *Female crime and delinquency: Critical perspectives and effective interventions* (pp. 283-322). Boston: Jones and Bartlett.

Brennan, T., Breitenbach, M., & Dieterich, W. (2008). Towards an explanatory taxonomy of adolescent delinquents: Identifying several social-psychological profiles. *Journal of Quantitative Criminology, 24*, 179-203.

Brennan, T., Dieterich, W., & Oliver, W. (2004). *The COMPAS scales: Normative data for males and females. Community and incarcerated samples.* Traverse City, MI: Northpointe Institute for Public Management.

Brennan, T., Dieterich, W., & Oliver, W. (2007). *COMPAS: Correctional offender management for alternative sanctioning. Technical manual and psychometric report* (V. 5.01). Traverse City, MI: Northpointe Institute for Public Management.

Brennan, T., Wells, D., & Alexander, J. (2004). *Enhancing prison classification systems: The emerging role of management information systems.* Washington, DC: U.S. Department of Justice, National Institute of Corrections.

Clements, C. B. (1996). Offender classification: Two decades of progress. *Criminal Justice and Behavior, 23,* 121-143.

Coleman, J. S. (1990). *Foundations of social theory*. Cambridge, MA: Harvard University Press.

Cooke, D. J., Forth, A. E., & Hare, R. D. (1996). *Psychopathy: Theory, research and implications for society.* Dordrecht, Netherlands: NATO Science Series.

Cullen, F., & Agnew, R. (2003). *Criminological theory: Past to present.* Los Angeles: Roxbury.

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.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision models. *American Psychologist, 34,* 571-582.

Elliott, D. S., Huizinga, D., & Ageton, S. S. (1985). *Explaining delinquency and drug use.* Beverly Hills, CA: Sage.

Ellis, L., & Walsh, A. (1997). Gene-based evolutionary theories in criminology. *Criminology, 35,* 229-276.

Estroff, S., Zimmer, C., Lachicotte, W., & Benoit, J. (1994). The influence of social networks and social support on violence by persons with serious mental illness. *Hospital and Community Psychiatry, 45,* 669-679.

Eysenck, S., & Eysenck, H. (1978). Impulsiveness and venturesomeness: Their position in a dimensional system of person-ality description. *Psychological Reports, 43,* 1247-1255.

Farr, K. A. (2000). Classification for female inmates: Moving forward. *Crime & Delinquency, 46,* 3-17.

Farrington, D. (1991). Childhood aggression and adult violence: Early precursors and later life outcomes. In D. Pepler & K. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 5-29). Hillsdale, NJ: Lawrence Erlbaum.

Farrington, D. P. (2003). Developmental and life course criminology: Key theoretical and empirical issues. *Criminology, 41,* 221-255.

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.

Farrington, D. P., & West, D. J. (1993). Criminal, penal, and life histories of chronic offenders: Risk and protective factors and early identification. *Criminal Behaviour and Mental Health, 3,* 492-523.

Fass, T. L., Heilbrun, K., DeMatteo, D., & Fretz, R. (2008). The LSI-R and the COMPAS: Validation on two risk-needs tools. *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), 44-48.

Gendreau, P., Little, T., & Goggin, C. (1996). A meta-analysis of the predictors of adult offender recidivism: What works! *Criminology, 34,* 575-607.

Glaser, D. (1987). Classification for risk. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making* (pp. 249-292). Chicago: University of Chicago Press.

Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime.* Stanford, CA: Stanford University Press.

Gottfredson, S. D. (1987). Prediction: An overview of selected methodological issues. In D. M. Gottfredson & M. Tonry (Eds.), *Prediction and classification: Criminal justice decision making* (pp. 21-52). Chicago: University of Chicago Press.

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.

Grove, W. M., & Meehl, P. E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy and Law, 2,* 293-323.

Hagen, J. (1998). Life course capitalization and adolescent behavioral development. In R. Jessor (Ed.), *New perspectives on adolescent risk behavior.* Cambridge, MA: Cambridge University Press.

Hannah-Moffatt, K., & Shaw, M. (2001). *Taking risks: Incorporating gender and culture into classification and assessment of federally sentenced women in Canada.* Ottawa, Ontario: Status of Women Canada.

Hardyman, P., & Van Voorhis, P. (2004). *Developing gender-specific classification systems for women offenders*. Washington, DC: U.S. Department of Justice, National Institute of Corrections.

Hare, R. D. (1991). *The Hare Psychopathy Checklist–Revised.* Toronto: Multi-Health Systems.

Harrell, F. E., Califf, R. M., Pryor, D. B., Lee, K. L., & Rosati, R. A. (1982). Evaluating the yield of medical tests. *Journal of the American Medical Association, 247,* 2543-2546.

Harris, P. W., & Jones, P. R. (1999). Differentiating delinquent youths for program planning and evaluation. *Criminal Justice and Behavior, 26,* 403-434.

Hastie, R., & Dawes, R. M. (2001). *Rational choice in an uncertain world: The psychology of judgment and decision-making*. Thousand Oaks, CA: Sage.

Hirschi, T. (1969). *Causes of delinquency.* Berkeley: University of California Press.

Hoffman, P. B. (1994). Twenty years of operational use of a risk prediction instrument: The United States Parole Commission's Salient Factor Score. *Journal of Criminal Justice, 22,* 477-494.

Holtfreter, K., & Cupp, R. (2007). Gender and risk assessment: The empirical status of the LSI-R for women. *Journal of Contemporary Criminal Justice, 23*, 363-382.

Jones, P. R. (1996). Risk prediction in criminal justice. In A. T. Harland (Ed.), *Choosing correctional options that work* (pp. 33-68). Thousand Oaks, CA: Sage.

Kalbfleisch, J. D., & Prentice, R. L. (2002). *The statistical analysis of failure time data* (2nd ed.). New York: John Wiley.

Kroner, D. G., Stadtland, M., Eidt, M., & Nedopil, N. (2007). The validity of the Violence Risk Appraisal Guide (VRAG) in predicting criminal recidivism. *Criminal Behaviour and Mental Health, 17,* 89-100.

Lilienfeld, S. O., & Andrews, B. P. (1996). Development and preliminary validation of a self-report measure of psychopathic personality traits in a non-criminal population. *Journal of Personality Assessment, 66,* 488-524.

Lipsey, M. W., & Derzon, J. H. (1998). Predictors of violent or serious delinquency in adolescence and early adulthood: A synthesis of longitudinal research. In R. Loeber & D. P. Farrington (Eds.), *Serious and violent juvenile offenders: Risk factors and successful interventions* (pp. 86-105). Thousand Oaks, CA: Sage.

Lösel, F. (1995). The efficacy of correctional treatment: A review and synthesis of meta-evaluations. In J. McGuire (Ed.), *What works: Reducing re-offending* (pp. 79-111). Chichester, UK: Wiley.

Lykken, D. T. (1995). *The antisocial personalities.* Mahwah, NJ: Lawrence Erlbaum.

Marcus, B. (2003). An empirical examination of the construct validity of two alternative self-control measures. *Educational and Psychological Measurement, 63,* 674-706.

Marris, P. (1987). *Loss and change.* New York: Pantheon.

McNaughton, C. (2007, September). *Life on the edge: Substance abuse and homelessness—Escape, resistance or deviance.* Paper presented at the annual meeting of the British Society of Criminology, Glasgow, UK.

Mealey, L. (1995). The sociobiology of sociopathy: An integrated evolutionary model. *Behavioral and Brain Sciences, 18,* 523-599.

Meier, S. (2002). *Bridging case conceptualization, assessment and intervention.* Thousand Oaks, CA: Sage.

Millon, T., & Davis, R. D. (1997). The place of assessment in clinical science. In T. Millon (Ed.), *The Millon inventories: Clinical and personality assessment* (pp. 3-22). New York: Guilford.

Mossman, D. (1994). Assessing predictions of violence: Being accurate about accuracy. *Journal of Consulting and Clinical Psychology, 62,* 783-792.

National Council on Crime and Delinquency. (2006). *Correctional Assessment and Intervention System.* Retrieved March 28, 2007, from http://www.nccd-crc.org/nccd/n_index_main.html

National Institute of Justice. (1999). *Annual report on drug use among adult and juvenile arrestees (1998). Arrestee Drug Abuse Monitoring Program (ADAM).* Washington, DC: Author.

National Research Council. (1993). *Understanding and preventing violence* (A. J. Reiss & J. A. Roth, Eds.). Washington, DC: National Academy Press.

Nelson, W. (1972). Theory and applications of hazard plotting for censored failure data. *Technometrics, 14,* 945-965.

Northpointe Institute for Public Management. (1996). COMPAS [Computer software]. Traverse City, MI: Author.

Osgood, D., Wayne, J. K., Wilson, J. G., Bachman, P., O'Malley, G., & Johnston, L. D. (1996). Routine activities and individual deviant behavior. *American Sociological Review, 61,* 635-655.

Palmer, T. (1992). *The re-emergence of correctional intervention.* Newbury Park, CA: Sage.

Parker, J., & Asher, S. (1987). Peer relations and later personal adjustment: Are low accepted children at risk? *Psychological Bulletin, 102,* 357-389.

Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (1998). *Violent offenders: Appraising and managing risk.* Washington, DC: American Psychological Association.

Reisig, M. D., Holtfreter, K., & Morash, M. (2006). Assessing recidivism risk across female pathways to crime. *Justice Quarterly, 23,* 384-405.

Salisbury, E. J., Van Voorhis, P., & Wright, E. (2006, November). *Construction and validation of a gender responsive risk/needs instrument for women offenders in Missouri and Maui.* Paper presented at the annual conference of the American Society of Criminology, Los Angeles.

Sampson, R. J., & Lauritsen, J. (1994). Deviant lifestyles proximity to crime and the offender-victim link in personal violence. *Journal of Research in Crime and Delinquency, 27,* 7-40.

Sampson, R. J., Raudenbush, S., & Earls, F. (1997). Neighborhoods and violent crime: A multilevel study of collective efficacy. *Science, 277,* 918-924.

Solomon, A., Johnson, K. D., Travis, J., & McBride, E. (2004). *From prison to work: The employment dimensions of prisoner reentry.* Washington, DC: Urban Institute.

Stalans, L. J., Yarnold, P. R., Seng, M., Olsen, D., & Repp, M. (2004). Identifying three types of violent offenders and predicting their recidivism and performance while on probation: A classification tree analysis. *Law and Human Behavior, 26,* 253-271.

Stevenson, H. C. (1998). Raising safe villages: Cultural-ecological factors that influence the emotional adjustment of adolescents. *Journal of Black Psychology, 24,* 44-59.

Tape, T. G. (2003). *Interpreting diagnostic tests: The area under the ROC curve.* Unpublished report, University of Nebraska Medical Center, Omaha.

Thornberry, T., Huizinga, D., & Loeber, R. (Eds.). (1995). The prevention of serious delinquency and violence: Implications from the program of research on the causes and correlates of delinquency. *Sourcebook on serious, violent and chronic juvenile offenders.* Thousand Oaks, CA: Sage.

Van Voorhis, P. (1994). *Psychological classification of the adult male prison inmate.* Albany: State University of New York Press.

Walters, G. D. (1995). The Psychological Inventory of Criminal Thinking Styles: Part I. Reliability and preliminary validity. *Criminal Justice and Behavior, 22,* 307-325.

Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. *Psychology, Crime and Law, 10,* 243-257.

Ward, T., & Stewart, C. (2003). Criminogenic needs and human needs: A theoretical model. *Psychology, Crime and Law, 9,* 125-143.

Warren, M. Q., & Hindelang, M. J. (1979). Differential explanation of offender behavior. In H. Toch (Ed.), *Psychology of crime and criminal justice* (pp. 166-182). Prospect Heights, IL: Waveland Press.

Widiger, T. A., & Lynum, D. R. (1998). Psychopathy and the five-factor model of personality. In T. Millon, E. Simonsen, M. Birkett-Smith, & R. Davis (Eds.), *Psychopathy: Antisocial, criminal and violent behavior* (pp. 171-186). New York: Guilford.

Wormith, J. S. (2001, July). Assessing offender assessment: Contributing to effective correctional treatment. *The ICCA Journal,* pp. 12-23.

Zhang, S., Farabee, D., & Roberts, R. (2007, October). *Predicting parolee risk of recidivism*. Paper presented at the 66th semi-annual meeting of the Association for Criminal Justice Research, Sacramento, CA.

Tim Brennan, PhD, is a senior research scientist at Northpointe Institute. His main research interests include risk assessment, pattern recognition, classification, and machine learning in the context of crime and delinquency.

William Dieterich, PhD, is director of research at Northpointe Institute. His research interests include developing and testing prognostic models for use in criminal justice agencies.

Beate Ehret, PhD, is a research analyst at Northpointe Institute. Her research interests include gender and crime, juvenile delinquency, and evidence-based practice in criminal justice.