

DEVELOPMENT AND PRELIMINARY VALIDATION OF A DEFENDANT AND OFFENDER SCREENING TOOL FOR PSYCHOPATHOLOGY IN INMATE POPULATIONS

CHRISTOPHER J. FERGUSON

Texas A&M International University

CHARLES NEGY

University of Central Florida

The development of instruments to assess psychopathology within inmate populations has lagged far behind the development of general clinical assessment measures. At present, few empirically validated assessment instruments are available, and forensic evaluations typically rely on clinical measures that were not designed for inmate populations. The present study details the development and initial validation of a new forensically relevant assessment tool designed for and tested on an inmate population. This instrument is designed to function as a screening assessment tool for use in forensic and correctional evaluations that may assist forensic evaluators in answering questions related to psychopathology more reliably and validly than tools currently in use. Psychometric properties and potential usefulness of the instrument are discussed.

Keywords: psychopathology; inmates; forensic evaluation; assessment; Defendant and Offender Screening Tool (DOST)

Instruments designed to assess psychopathology within forensic and correctional populations are surprisingly limited given the large number of individuals arrested and incarcerated within the United

AUTHORS' NOTE: *Correspondence concerning this article should be addressed to Christopher J. Ferguson, Department of Behavioral, Applied Sciences and Criminal Justice, Texas A&M International University, 5201 University Blvd, Laredo, TX 78041; phone: (956) 326-2636; e-mail: CJFerguson1111@aol.com.*

CRIMINAL JUSTICE AND BEHAVIOR, Vol. 33, No. 3, June 2006 325-346

DOI:10.1177/0093854806286194

© 2006 American Association for Correctional and Forensic Psychology

States. Although the testimony of psychologists has been admissible since 1962 (*Jenkins v. United States*, 1962), Otto and Heilbrun (2002) suggest that forensic psychologists lack clear standards for evaluation and that psychometrically validated measures specifically developed for inmate populations remain largely unavailable. Although some progress had been made regarding the development of forensic psychology (that is, the subspecialty of clinical psychology concerned with assessment issues that pertain to legal and/or correctional settings), forensic evaluation as a clinical science has remained in its infancy. Despite the fact that the number of evaluation tools specifically designed for use in forensic assessments has increased, concerns regarding the psychometric properties of many of these tools remain. As a result, the utility of many of these measures as well as their ability to meet judicial standards, such as those set forth by *Daubert v. Merrell Dow Pharmaceuticals, Inc.* (1993) or *Frye v. United States* (1923), are limited.

The National Institute of Corrections has recently instituted reforms for assessment instruments used to classify inmates (see Brennan, 1987, for a full discussion) in an attempt to promote the development of empirically derived assessment instruments for inmates. Several measures stand out as forensic instruments that have demonstrated adequate psychometric properties, including, most notably, the Psychopathy Checklist-Revised (PCL-R; Hare, 1991). However, the PCL-R itself requires extensive training to administer, making its use prohibitive to many correctional and forensic institutions. Furthermore, the Historical/Clinical/Risk Management (Webster, Douglas, Eaves, & Hart, 1997) and Violence Risk Appraisal Guide (Webster, Harris, Rice, Cormier, & Quinsey, 1994) have been developed as actuarial risk for violence measures, although some controversy over their ability to predict violence persists (see, for example, Steadman et al., 2000, for a full discussion). Regarding competency evaluations, the MacArthur Competence Assessment Tool (Otto et al., 1998) shows promise in providing information relevant to such decisions. The Carlson Psychological Survey (Carlson, 1981) has been developed as a self-report screening tool for psychopathology and substance abuse in criminal populations. This measure has demonstrated adequate reliability with both adult (Carlson, 1981) and adolescent (Wright & Friesen, 1985) populations. Similarly, the Referral

Decision Scale (Teplin & Schwartz, 1989), a brief 14-item measure designed as a screener for mental illness in individuals entering correctional settings, is available. However, several studies have indicated problems with both this measure's validity and usability (e.g., Rogers, Sewell, Ustadt, & Reinhardt, 1995).

Aside from several of the measures discussed above, forensic measures of serious pathology remain in short supply. Thus, forensic evaluators often turn to measures developed for general clinical populations that have been applied to inmate populations (see, Otto & Heilbrun, 2002, for a full review). This class of assessment instruments includes the Minnesota Multiphasic Personality Inventory-2 (MMPI-2; Hathaway & McKinley, 1989), Personality Assessment Inventory (Morey, 1991), and Millon Clinical Multiaxial Inventory-III (Millon, 1994). These instruments have benefited from a wealth of research and generally demonstrate adequate psychometric properties. However, as argued by Otto and Heilbrun (2002), these measures were not originally designed for an offender and defendant population and were therefore not originally normed on such. They also are extensive, time-consuming instruments to complete and score. Moreover, Podboy and Kastl (1993), as well as Rogers (2003), suggest that the potential for misuse of clinical measures in forensic settings is high because of the unsuitability of such measures for forensic evaluations. These measures also leave a great deal of room for individual psychologists' interpretations of the test results, as such results require subsequent translation into legally relevant observations. Research on the validity of briefer screening measures for psychopathology, such as the Brief Symptom Inventory (Overall, Henry, & Markett, 1972) or Symptoms Checklist-90 (Derogatis, 1992), has been sparse and indicated only modest predictive and diagnostic utility (see, e.g., Ferguson et al., in press; Wilson, Taylor, & Robertson, 1985).

It is not sufficient to simply renorm existing clinical scales on inmate populations, as the scales themselves may fail to tap into the constructs of interest. For example, items that adequately distinguish between normal and psychotic individuals in noninmate populations may fail to differentiate adequately with inmate populations. As such, the scales themselves may lose meaning when transferred to a new population for which they were not intended. For example, as

noted in Brennan (1987), despite systematic attempts to use the MMPI-2 to detect psychopathology in correctional settings, its ability to function well at this task remains doubtful. This finding was further highlighted by a study examining the MMPI-2 profiles of 161 offenders, in which both validity and clinical elevations were commonly noted, even in the absence of a specific mental disorder (Wright, Nussbaum, Lynett, & Buis, 1997).

PRESENT STUDY

The goal of the present study was to develop and provide initial support for an assessment instrument that can be used as a relatively brief screening tool for a variety of forensically relevant constructs. The current instrument was designed to provide information in forensic evaluations as well as to be used as a screener for serious mental or cognitive impairment in correctional populations. Reviews (e.g., Johnstone, Schopp, & Shigaki, 2000; Meloy & Gacono, 2000; Otto & Heilbrun, 2002) suggest that the most salient issues for forensic assessment fall into the following categories:

1. validity issues, including both social desirability and malingering, and
2. forensically relevant scales, including scales that provide information related to insanity/competency (specifically, psychosis and mental retardation) as well as violence propensity/aggressiveness.

In the present study, initial development and psychometric support for the Defendant and Offender Screening Tool (DOST) are described. The following goals were pursued in this study:

1. create theoretically derived items that adhered to *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 1994) criteria for psychopathology, or which otherwise made conceptual sense in a true/false format for each of the five related subscales (Social Desirability, Malingering, Psychosis, Cognitive Impairment, and Aggressiveness);
2. establish internal consistency and test-retest reliability for each subscale; and
3. establish initial support for construct validity on each subscale based on the present sample of jail inmates and noninmate participants.

TABLE 1: Involvement of Participants in Each Study Phase

	<i>Phase 1 (Subscale Only)</i>	<i>Phase 2 (Whole Test)</i>
Social Desirability	40 control inmates 40 experimental inmates	36 control inmates 35 experimental inmates
Malingering	40 malingering inmates 44 schizophrenic inmates	37 malingering inmates 35 schizophrenic inmates
Psychosis	45 control inmates 44 schizophrenic inmates	34 control inmates 35 schizophrenic inmates
Cognitive Impairment	30 college students	44 inmates
Aggressiveness	41 nonviolent inmates 43 violent inmates	34 nonviolent inmates 36 violent inmates
Factor Analysis	381 inmates (including those described above)	

METHOD

PARTICIPANTS

Participants for the present study included 674 inmates recruited from a county jail located in Florida as well as 30 university students who served as participants for the initial item construction of the Cognitive Impairment subscale (a brief test of verbal intelligence). Involvement of participants in the two study phases is presented in Table 1.

Of the 30 university students, 6 (20%) were male and the remaining 24 (80%) were female. Regarding ethnicity, 22 (73.3%) were Caucasian, 2 (6.7%) were African American, and 6 (20%) were Hispanic. Their mean age was 27 ($SD = 10.46$). All student participants were offered extra credit in exchange for their participation in this project. Of the 674 inmates, 452 (67.1%) were male and 222 (32.9%) were female. Regarding ethnicity, 292 (43.3%) were Caucasian, 292 (43.3%) were African American, 74 (11%) were Hispanic, 4 (0.6%) were Asian, and 11 (1.6%) did not fall into one of the four previous categories. Regarding the nature of their charges, 147 (21.8%) were charged with violent felonies (e.g., aggravated assault involving aggressive or assault behavior toward another person), 16 (2.4%) with sexual assaults, 338 (50.1%) with nonviolent felonies, 163 (24.2%) with misdemeanors, and 4 (0.5%) did not have information available on their charges at the time they were interviewed. The average age of

inmate participants was 34.2 ($SD = 10.06$), and they had an average of 12 years of education ($SD = 2.74$ years).

Inmates were selected randomly from among those who were available at the jail on the days this study took place. In some cases, specific inmates were selected because of the nature of their diagnosis and the relative rarity of their condition within the jail population (e.g., inmates with a diagnosis of schizophrenia, schizoaffective disorder, or psychotic disorder not otherwise specified). All inmates were approached individually and invited to participate in the study after reading, or having read to them, an informed consent and confidentiality agreement, as well as being invited to ask any questions they had. The examiner judged the ability of individual participants to give informed consent before participation in the study. Of the original 744 inmates who were approached for participation in the study, 70 (9.4%) exercised their right to refuse. Informed consent procedures conformed to both university and jail internal review board standards as well as to American Psychological Association standards for research on protected populations. No information was shared with jail personnel, and research materials were not stored on jail premises. In accordance with jail regulations, inmates were not offered compensation for their participation.

PROCEDURE

Initial item construction employed a combination of rational and empirical methods. All items were designed to apply to forensic and correctional settings and were designed with inmate participants in mind. Content validity of the subscales was demonstrated in several ways. First, clear content domains for each subscale were described, and items were written to correspond to those specific content domains. Trained assistants (including both licensed clinical psychologists working in a correctional setting as well as clinical psychology graduate students) familiar with forensic evaluation procedures and standards and familiar with the relevant content domains conducted initial development of items for the five subscales of the DOST. An initial pool of between 25 and 56 true/false items was developed for each subscale (the exact number of initial items per subscale varied). This pool of items was then scrutinized for content and adherence to

the underlying theoretical constructs by a second independent body of reviewers, including two jail psychologists involved in the design and selection of clinical screening measures for a correctional facility. Items that were objected to by any reviewer were discarded. The content validity of the resultant scales was further confirmed by examining item-total correlations. Although item-total correlations generally reflect reliability, they also reflect content grouping (Lachar & Gruber, 1995). The resultant item-total correlations are presented in Table 2.

Each of the items on the subscales was then evaluated for individual utility (Phase 1). In the case of most subscales (except for cognitive impairment), this process consisted of administration of the total item subscale to a sample of inmates targeted by the specific intent of the subscale (e.g., violent felons or mentally ill inmates) and a comparison sample (inmates from the general inmate population). In all cases, the questions were read aloud to the inmates to minimize the effect of reading problems. Given that the DOST may be used on inmates with cognitive impairment, chronic mental illness, or reading difficulties, adopting this procedure appeared to be pertinent. Using the SMOG formula (McLaughlin, 1969), the reading level of the DOST was calculated to be at the sixth-grade level. Phi-coefficient correlations were conducted for the scores of each of the individual items with sample group membership (experimental or comparison). Items that did not correlate with group membership at a statistically significant (i.e., $p \leq .05$) level were discarded.

In the case of the Cognitive Impairment subscale, the items were administered to college students along with the Verbal scale of the Wechsler Adult Intelligence Scale-III (WAIS-III; Wechsler, 1997). Because an intelligence level in the mentally retarded range may be relevant for competency, the cognitive impairment scale was designed to briefly screen for individuals with potential mental retardation. Although this scale, such as the others, was ultimately intended for an inmate population, using a nonforensic sample in the initial development of items was deemed necessary to prevent the potential for a ceiling effect and to provide some cross-validation. Using college students for Phase 1 was intended to screen out items that had little or no correlation with verbal intelligence. The items that were retained were subsequently validated on an inmate sample in Phase 2.

TABLE 2: Individual Item to Total Scale Scores on the DOST Subscales

Social Desirability		Malingering		Psychosis		Cognitive Impairment		Aggressiveness	
Item #	Correlation	Item #	Correlation	Item #	Correlation	Item #	Correlation	Item #	Correlation
1	.60	1	.51	1	.57	1	.25	1	.59
2	.38	2	.74	2	.64	2	.54	2	.61
3	.65	3	.75	3	.63	3	.45	3	.38
4	.60	4	.82	4	.68	4	.40	4	.44
5	.75	5	.66	5	.73	5	.46	5	.47
6	.47	6	.74	6	.72	6	.55	6	.54
7	.70	7	.66	7	.55	7	.52	7	.46
8	.39	8	.80	8	.47	8	.47	8	.59
9	.42	9	.65	9	.70	9	.52	9	.63
10	.60	10	.63	10	.57	10	.61	10	.50
11	.56	11	.57	11	.59	11	.37	11	.58
		12	.73	12	.57	12	.50	12	.61
				13	.59	13	.42		
				14	.64				
				15	.65				
				16	.73				
				17	.61				
				18	.63				
				19	.62				

Note. DOST = Defendant and Offender Screening Tool.

Individual items of the Cognitive Impairment subscale were correlated with the total WAIS-III Verbal scale.

The resultant subscales then were given to a second group of inmates to test for reliability and validity (Phase 2). As during Phase 1, each of the resultant subscales were given to a target sample and a comparison sample. Each subscale was assessed for Cronbach's alpha internal consistency and its ability to discriminate between a target and comparison sample (criterion-related validity). In the case of the Cognitive Impairment subscale, the total subscale score was correlated with the WAIS-III Verbal scale with a sample of inmates. Test-retest reliability for all five subscales was tested using a separate group of 30 inmates. The retest was administered within 7 to 10 days of the original administration. Although a longer test-retest period would have been desired, this was found to be impractical given the relatively high turnover rate of jail inmates.

The results of the total DOST scale for the 381 inmates involved in Phase 2 were further analyzed using a principal components factor analysis to determine if the items on the subscales loaded adequately on each of the subscales with which they were designed to function. As noted by MacCallum (1998), it is unlikely that factors within any given factor structure will be totally uncorrelated, and even if they are uncorrelated, the factor structure will remain evident with an oblique rotation. As such, a promax oblique rotation was employed during the factor analysis. Example items from each of the DOST subscales are presented in the appendix.

All inmates in comparison groups as well as experimental groups for the Social Desirability and Malingering subscales were randomly assigned.

RESULTS

PHASE 1: INITIAL ITEM CONSTRUCTION

Social Desirability. The Social Desirability subscale was designed to test for responses marked by defensiveness or denial of normal but embarrassing human qualities, including normative aggressive responses or minor law-breaking behaviors. The Social

Desirability subscale was administered to two samples of inmates. The first group of 40 inmates was told to take the test presenting themselves as if they felt pressured to present themselves as good as possible. Specifically, the inmates were instructed to take the test "as if you have no psychological or personal problems or flaws whatsoever." The second group of 40 inmates was not given any special directions. Phi-coefficient correlations were computed between each individual item and group membership. Of the original pool of 25 items, 11 items that correlated with group membership at the $p = .01$ level were retained for Phase 2. The resultant 11-item subscale demonstrated a Cronbach's alpha coefficient of .87 in Phase 1.

Malingering. The Malingering subscale was designed to test for a response style marked by falsification of mental health symptoms. This subscale was administered to a sample of 44 inmates formally diagnosed by jail psychologists or psychiatrists with a psychotic disorder as well as to a sample of 40 inmates without psychosis. The diagnosis of a psychotic disorder had been obtained using *DSM-IV* criteria and was reconfirmed by research staff using the psychosis section of the Structured Clinical Interview for the *DSM-IV* (SCID; First, Spitzer, Gibbon, & Williams, 1997). Inmates in the control group denied experiencing *DSM-IV* criteria for schizophrenia or other psychotic disorders. These inmates were thus instructed to predict how they thought a mentally ill person would respond to the items and to respond accordingly rather than responding how they actually felt themselves. Specifically, nonpsychotic inmates were instructed to respond "as if you were really crazy." Inmates with schizophrenia were simply instructed to respond to the items as honestly as possible. This procedure facilitated the testing for items that distinguished between individuals diagnosed with schizophrenia versus how nonschizophrenic inmates imagined people with schizophrenia might respond to such items. Phi-coefficient correlations were computed between each individual item and group membership. Of the original pool of 31 items, 12 items that correlated at the $p = .01$ level were initially retained for Phase 2. The resultant 12-item subscale demonstrated a Cronbach's alpha coefficient of .88 during Phase 1.

Psychosis. The Psychosis subscale was designed to measure symptoms or behaviors consistent with a psychotic disorder in accordance with *DSM-IV*. The Psychosis subscale was administered to a sample of 44 inmates diagnosed by jail psychologists or psychiatrists with a psychotic disorder and to a sample of 45 inmates without schizophrenia who were instructed to answer normally. Diagnosis of schizophrenia or related psychotic disorder was reconfirmed using the SCID. That inmates in the control group did not have symptoms of schizophrenia was also confirmed. Nonetheless, it should be noted that the possibility remains for false negatives in the control group should some inmates have not endorsed psychotic symptoms they possessed. Phi-coefficient correlations were computed between each individual item and group membership. Of the original pool of 40 items, 19 items that correlated at the $p = .01$ level were retained for Phase 2. The resultant 19-item subscale demonstrated a Cronbach alpha coefficient of .93 during Phase 1.

Cognitive Impairment. The Cognitive Impairment subscale was intended to serve as a brief screening tool to identify inmates who likely have impaired cognitive abilities and therefore would warrant a more thorough cognitive assessment for confirmation. The Cognitive Impairment subscale was administered to a college sample ($n = 30$) along with the WAIS-III Verbal scale. Scores on the WAIS-III Verbal scale within this sample ranged from 84 to 128 ($M = 105$, $SD = 11.8$). Point-biserial correlations were computed between individual items of the Cognitive Impairment subscale and Verbal scale scores of the WAIS-III. The Verbal scale score was chosen because of its ability to provide reliable and accurate assessments even when performance scales are not used (Taylor & Heaton, 2001). Of the original pool of 47 items, 13 items that significantly correlated or whose correlation approached significance with the WAIS-III Verbal scale were retained for Phase 2. As the relatively small sample size ($n = 30$) limited the power of these analyses, several items that approached but did not reach statistical significance were retained to maintain enough items for a stable subscale. The resultant 13-item subscale demonstrated a Cronbach's alpha coefficient of .70 during Phase 1 and correlated $-.66$ ($p = .01$) with

the WAIS-III Verbal scale among this sample of undergraduate students. A negative correlation was the expected direction given that a higher level of cognitive impairment theoretically ought to be associated with a lower verbal IQ score.

Aggressiveness. The Aggressiveness subscale was designed to measure antisocial tendencies, particularly those that might distinguish individuals at high risk for aggressive or violent behavior apart from other individuals. The Aggressiveness subscale was administered to a sample of 43 inmates with a record of arrests for violent felonies (a history of at least two felonies, such as aggravated assault, involving aggressive or assault behavior toward another person) and to a comparison sample of 41 inmates with no history of violence. It bears mentioning that the presence or absence of a history of violent felonies is not a perfect indicator of all violent behavior. Phi-coefficient correlations were computed between each individual item and group membership. Of the original 56 items, 12 items that correlated at the $p = .01$ level were retained for Phase 2. The resultant 12-item subscale demonstrated a Cronbach's alpha coefficient of .81 during Phase 1.

PHASE 2: RELIABILITY AND VALIDITY

The analyses conducted during Phase 1 resulted in a 67-item questionnaire that was administered to the second group of 381 inmates. Target and control groups were similar to those in Phase 1. As no differences were found in the scores of male and female inmates, the two groups were condensed and analyzed for reliability and validity together. Results of the reliability and validity analyses reported in this section are summarized in Table 3.

Social Desirability. The validity of the Social Desirability subscale was examined with 71 inmates, who were assigned randomly to one of two conditions: 35 inmates in a target group and comparison group of 36 inmates. The results of the target and comparison groups were compared using an independent samples t test. Results indicated that the Social Desirability subscale was able to significantly discriminate between defensive ($M = 9.74$, $SD = 1.56$) and

TABLE 3: Results of Reliability and Validity Analyses in Phase 2

<i>Subscale</i>	<i>Coefficient Alpha</i>	<i>Test-Retest</i>	<i>t</i>	<i>df</i>
Social Desirability	.82	.75	13.29*	58 ^a
Malingering	.89	.71	16.38*	70
Psychosis	.91	.94	11.85*	45 ^a
Cognitive Impairment	.70	.79	3.94*	42
Aggressiveness	.77	.90	12.24*	42 ^a

a. In the case of this analysis, Levine's test for equality of variances indicated that equal variances could not be assumed and the degrees of freedom reflect appropriate adjustments to the *t* test analysis.

* $p \leq .01$.

normal ($M = 3.17$, $SD = 2.51$) respondents, $t(58) = 13.29$, $p = .01$. Individual item to total score correlations for all subscales are presented in Table 2.

Malingering. The validity of the Malingering subscale was examined with 72 inmates assigned to one of two conditions: 37 inmates in a target falsification group and 35 inmates with schizophrenia in a comparison group. The results of these two groups were compared using an independent samples *t* test. Results indicated that the Malingering subscale was able to significantly discriminate between malingering ($M = 9.81$, $SD = 2.22$) and actual schizophrenic ($M = 2.34$, $SD = 1.57$) responders, $t(70) = 16.38$, $p = .01$.

Psychosis. The validity of the Psychosis subscale was examined with 69 inmates, who were assigned to one of two conditions: 35 inmates who had been diagnosed with schizophrenia by a jail psychologist or psychiatrist and a comparison group of 34 inmates who had not been diagnosed with schizophrenia. Results indicated that the Psychosis subscale significantly discriminated between schizophrenic ($M = 12.63$, $SD = 4.65$) and nonschizophrenic ($M = 2.56$, $SD = 1.88$) inmates, $t(45) = 11.85$, $p = .01$. Individual item to total score correlations are presented in Table 2.

Cognitive Impairment. The validity of the Cognitive Impairment subscale was examined with 44 inmates, who were also administered the WAIS-III Verbal scale. Scores on the WAIS-III Verbal scale

with this sample ranged from 61 to 137 ($M = 82$, $SD = 16.2$). Of these inmates, 16 (36%) tested with verbal IQs lower than 70, which allowed for testing a range of IQ scores, including those that fell in the impaired range. The Cognitive Impairment subscale was correlated with the WAIS-III Verbal scale at $r = -.78$. Correlations with WAIS-III Verbal subscales ranged between $-.33$ and $-.80$. The scores of the normal (verbal IQ > 70 , $n = 28$) and mentally retarded (verbal IQ < 70 , $n = 16$) inmates on the Cognitive Impairment subscale were compared using an independent samples t test. Results indicated that the Cognitive Impairment subscale was able to significantly discriminate between mentally retarded inmates ($M = 9.06$, $SD = 2.02$) and normal inmates ($M = 6.29$, $SD = 2.37$), $t(42) = 3.94$, $p = .01$. It should be noted that the terms *normal* and *mentally retarded* are not used here as diagnostic categories but rather as broad descriptors of performance range on the WAIS-III Verbal scale.

Aggressiveness. The validity of the Aggressiveness subscale was examined with 70 inmates, who were assigned to one of two conditions: 36 inmates in a violent arrest history target group and a comparison group of 34 inmates. The results of the target and comparison groups were compared using an independent samples t test. Results indicated that the Aggressiveness subscale significantly discriminated between inmates with ($M = 6.31$, $SD = 2.66$) and inmates without ($M = 0.59$, $SD = 0.86$) a history of violent crimes, $t(42) = 12.24$, $p = .01$.

SENSITIVITY AND SPECIFICITY

To assess scores with the maximum clinical utility, the sensitivity and specificity of clinical cutoff scores were assessed. Similarly, Area Under the Curve Receiver Operating Characteristic analyses were conducted on the five DOST subscales. The five subscales of the DOST were found to have adequate to good sensitivity and specificity within the current sample. Sensitivity and specificity data for each of the five scales, as well as clinical cutoff scores, are presented in Table 4.

FACTOR ANALYSIS

A principal components factor analysis was conducted on the DOST data obtained from 381 inmates to examine the construct

TABLE 4: Sensitivity, Specificity, Area Under the Curve, and Clinical Cutoff for Five DOST Subscales

<i>Subscale</i>	<i>Sensitivity</i>	<i>Specificity</i>	<i>ROC</i>	<i>95% Confidence Interval</i>	<i>Clinical Cutoff</i>
Social Desirability	.94	0.92	.97*	0.94 to 1.00	8
Malingering	.95	0.94	.98*	0.97 to 1.00	5
Psychosis	.86	1.00	.96*	0.90 to 1.00	9
Cognitive Impairment	.88	0.80	.92*	0.84 to 1.00	7
Aggressiveness	.78	1.00	.97*	0.93 to 1.00	5

Note. ROC = Area Under the Curve Receiver Operating Characteristic; DOST = Defendant and Offender Screening Tool.

* $p \leq .001$.

validity of the scales. It was anticipated that individual items would load on their representative factors at least .30. Initial results suggested that the five factors accounted for 42% of the variance. As expected, each of the items generally loaded with other items from the same subscale, although several items either loaded slightly less than .30 or loaded on two separate factors (although each item did load best on the theoretically relevant factor). Rather than necessarily indicating poor items, it could be possible that these items may ultimately identify specific sorts of risks. For example, items that load best on the Aggressiveness subscale but also load on the Psychosis subscale may be best at predicting aggressiveness in individuals with schizophrenia or other psychotic illnesses. Table 5 presents the loadings of individual items on each of the five factors. Table 6 presents intercorrelations between the DOST subscales.

DISCUSSION

The DOST is relatively brief (67 items; approximate 15-min administration time) and is intended to serve as a screening tool to red-flag individuals about to stand trial or commence incarceration who may need a more comprehensive evaluation. The results of this study provide initial support for the reliability and validity of the DOST as a forensically relevant measure of psychopathology. Specifically, all subscales of the DOST demonstrated acceptable coefficient alpha internal consistencies and test-retest reliabilities.

TABLE 5: Individual Item Loadings on Five-Factor Model

<i>Item #</i>	<i>Psychosis</i>	<i>Malingering</i>	<i>Social Desirability</i>	<i>Cognitive Impairment</i>	<i>Aggressiveness</i>
Soc1	.15	-.26	.53	-.14	.09
Soc2	.00	-.13	.46	.05	-.12
Soc3	.11	-.29	.50	-.08	.30
Soc4	.03	-.32	.46	-.05	.30
Soc5	.10	-.32	.60	-.03	.37
Soc6	.06	-.03	.60	.11	.00
Soc7	.17	-.36	.49	-.03	.40
Soc8	-.05	-.08	.44	.13	.07
Soc9	.06	-.01	.55	.01	-.02
Soc10	.15	-.25	.50	-.11	.26
Soc11	-.05	-.31	.44	.12	.31
Mal1	-.15	.40	-.24	.03	.10
Mal2	-.06	.73	-.15	.07	.09
Mal3	.08	.775	-.07	.10	.17
Mal4	.09	.82	-.10	.19	.08
Mal5	-.03	.64	-.14	.05	.02
Mal6	.14	.75	-.07	.12	.05
Mal7	-.04	.69	-.17	.11	-.10
Mal8	.07	.67	-.10	.05	.12
Mal9	.08	.89	-.01	.10	.03
Mal10	-.11	.68	.02	.03	-.13
Mal11	.06	.61	-.11	.08	.09
Mal12	-.38	.49	-.29	.04	-.10
Psy1	.61	-.04	-.11	.02	.10
Psy2	.66	-.03	-.05	.22	.09
Psy3	.64	-.09	-.02	.14	.10
Psy4	.67	-.07	-.06	.22	.14
Psy5	.73	.01	.11	.14	.13
Psy6	.72	.02	.15	.17	.14
Psy7	.54	.10	.14	-.02	.28
Psy8	.45	.20	.15	.18	.26
Psy9	.69	.04	.20	.18	.08
Psy10	.53	.16	.25	.12	.24
Psy11	.60	-.09	-.11	.05	.14
Psy12	.58	.03	.14	.17	.06
Psy13	.60	.08	.23	.05	.09
Psy14	.64	-.07	-.07	.21	.14
Psy15	.66	-.08	-.08	.08	.15
Psy16	.71	.03	.15	.07	.23
Psy17	.63	.10	.10	.03	.11
Psy18	.68	.01	.04	-.04	.07
Psy19	.61	-.09	-.02	.16	.16

(continued)

TABLE 5 (continued)

<i>Item #</i>	<i>Psychosis</i>	<i>Malingering</i>	<i>Social Desirability</i>	<i>Cognitive Impairment</i>	<i>Aggressiveness</i>
Cog1	.17	.10	-.01	.26	-.01
Cog2	.23	.02	-.05	.56	.01
Cog3	.12	.28	.22	.53	.13
Cog4	-.01	-.04	-.22	.52	.09
Cog5	.18	-.08	-.22	.50	.23
Cog6	.07	.09	.07	.62	.18
Cog7	.12	.17	.26	.45	-.02
Cog8	.13	.02	-.03	.55	-.05
Cog9	.08	.09	.06	.58	.19
Cog10	.16	.04	-.04	.67	.16
Cog11	.07	.29	.23	.49	-.09
Cog12	.11	.19	.21	.53	.14
Cog13	.10	.06	.01	.50	-.01
Agg1	.32	.11	.29	.14	.46
Agg2	.20	.02	.03	.00	.64
Agg3	.10	.04	-.02	.16	.38
Agg4	.09	.07	-.05	.13	.48
Agg5	.05	.07	.10	.08	.51
Agg6	.25	.04	.05	.05	.52
Agg7	.24	.13	.06	.20	.37
Agg8	.41	-.09	-.02	.12	.49
Agg9	.29	.02	.12	.01	.39
Agg10	.31	.02	.13	-.04	.48
Agg11	.15	-.01	.11	.04	.66
Agg12	.42	.11	.30	.09	.47

Note. Soc = Social Desirability; Mal = Malingering; Psy = Psychosis; Cog = Cognitive Impairment; Agg = Aggressiveness.

TABLE 6: Intercorrelations Between DOST Subscale

<i>Subscale</i>	<i>Agg</i>	<i>Psy</i>	<i>Cog</i>	<i>Mal</i>	<i>Soc</i>
Aggressiveness (Agg)		.52	.12	.49	-.57
Psychosis (Psy)			.18	.56	-.36
Cognitive Impairment (Cog)				.18	-.05
Malingering (Mal)					-.24
Social Desirability (Soc)					

Note. DOST = Defendant and Offender Screening Tool.

Furthermore, all subscales of the DOST demonstrated the ability to discriminate between a target and comparison group. The Cognitive Impairment subscale further demonstrated good concurrent

validity with the WAIS-III Verbal scale. Some initial support for construct validity of the subscales was supported by factor analysis, which demonstrated that each of the items loaded as expected on the appropriate theoretical factor. Taken together, these results provide preliminary support for use of the DOST in forensic and correctional settings, although much follow-up research is needed.

Some discussion of the interscale correlation between the Malingering subscale and Psychosis subscale is warranted. High scores on the Malingering subscale are designed to reflect a test-taking style of normal individuals who are attempting to fake mental illness. However, as these questions also involve descriptions of severe mental symptoms, it was not surprising to observe some correlation ($r = .56$) between these subscales. This finding is in keeping with the observation that the MMPI-2 Schizophrenia and F subscales are correlated at $r = .69$ (Hathaway & McKinley, 1989). For the Malingering subscale to function as desired, it would be expected that normal respondents, who are instructed to simply respond to the items, would endorse fewer items on the Malingering subscale than schizophrenic responders, who in turn should endorse fewer items than nonschizophrenic responders who are instructed to complete the items while purposefully trying to fake mental illness in their responses. In the present study, this is, in fact, what was observed. Normal respondents on the Malingering subscale scored a mean of 1.19 ($SD = 1.38$), whereas inmates with schizophrenia scored a mean of 2.34 ($SD = 1.57$), and nonschizophrenic inmates purposefully attempting to falsify a mental illness scored a mean of 9.81 ($SD = 2.22$).

FUTURE RESEARCH

In regard to future research projects, several would appear to be particularly useful to further examine the validity and potential usefulness of the DOST in multiple correctional settings. Although the DOST is expected to be generalizable across inmate groups (minimum security, maximum security, white collar, and so forth), only further study will demonstrate this. It would be specifically useful to

examine the utility of the DOST with prison samples, probationers, and pretrial versus posttrial inmates, as well as samples across the minimum- to maximum-security spectrum.

Concurrent validity studies conducted with the DOST would be helpful in further establishing its validity. At present, only the Cognitive Impairment subscale has been assessed for concurrent validity. However, testing concurrent validity of the other subscales will require them to be correlated with scales on measures on which the DOST was designed to improve. The most evident examples are the K, F, Schizophrenia, and Psychopathic Deviate scales on the MMPI-2, although other clinical measures or ratings may also be used. Although it is expected that the DOST will correlate with those MMPI-2 subscales, some differences between the DOST and the MMPI-2 (and other measures) are expected given the specific design of the DOST for use with inmate populations. Similarly, for the Malingering subscale, the Miller Forensic Assessment of Symptoms Test (Miller, 2004) may make a suitable instrument for examining concurrent validity. The Malingering subscale would also benefit from further analysis using known malingerers as a comparison group.

The current version of the DOST was designed to function primarily as a screener for serious mental illness and cognitive impairment as well as aggressiveness. The DOST is not designed to screen for other issues, such as suicide risk or substance abuse. Should the need for assessment materials in these domains continue to be unmet in correctional facilities, further research could be directed at examining the potential of the DOST for evaluating these constructs, including creating new subscales as necessary.

In conclusion, the DOST has performed satisfactorily in its initial evaluation, although continued research is necessary to fully demonstrate its potential and limitations. Much follow-up work is needed to confirm the utility of the DOST as a brief screening instrument for severe psychopathology in inmate populations. It is hoped that further research on the DOST will continue to demonstrate its utility across diverse populations and correctional settings.

APPENDIX

Defendant and Offender Screening Tool Example Items

Social Desirability:

If I could tell a lie to get out of trouble I would. (F)

Some people make me so mad that I want to hit them. (F)

Malingering:

I am so crazy I can't control my actions. (T)

People can make themselves happier if they want to. (F)

Cognitive Impairment:

Abraham Lincoln was the first president of the United States. (T)

A Tango is a type of fruit. (F)

Psychosis:

Other people or spirits are speaking directly into my brain. (T)

I am worried that other people can read my thoughts. (T)

Aggressiveness:

It disgusts me when I see other people crying. (T)

Most people are idiots. (T)

REFERENCES

- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Brennan, T. (1987). Classification: An overview of selected methodological issues. *Crime and Justice*, 9, 201-248.
- Carlson, K. A. (1981). A modern personality test for offenders: The Carlson Psychological Survey. *Criminal Justice and Behavior*, 8, 185-200.
- Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).
- Derogatis, L. (1992). *SCL-90-R: Administration, scoring & procedures manual-II for the revised version and other instruments of the psychopathology rating scale series*. Baltimore, MD: Clinical Psychometric Research.
- Ferguson, C. J., Averill, P. M., Rhoades, H., Rocha, D., Gruber, N., & Gummattira, P. (in press). Social isolation, impulsivity and depression as predictors of aggression in a psychiatric inpatient population. *Psychiatric Quarterly*.
- First, M., Spitzer, R., Gibbon M., & Williams, J. (1997). *Structured Clinical Interview for DSM-IV personality disorders, (SCID-II)*. Washington, DC: American Psychiatric Press.
- Frye v. United States, 293 F. 1013 (DC Cir. 1923).
- Hare, R. (1991). *The manual for the revised Psychopathy Checklist*. Toronto, Ontario, Canada: Multi-Health Systems.
- Hathaway, S., & McKinley, J. (1989). *MMPI-2: Manual for administration and scoring*. Minneapolis: University of Minnesota Press.

- Jenkins v. United States, 307 F.2d 637 (DC Cir. 1962).
- Johnstone, B., Schopp, L., & Shigaki, C. (2000). Forensic psychological evaluation. In R. Frank & T. Elliott (Eds.), *Handbook of rehabilitation psychology* (pp. 345-358). Washington, DC: American Psychological Association.
- Lachar, D., & Gruber, C. P. (1995). *Personality Inventory for Youth (PIY) manual*. Los Angeles: Western Psychological Services.
- MacCallum, R. (1998). Commentary on quantitative methods in I-O research. *The Industrial Organizational Psychologist*, 35, 19-30.
- McLaughlin, G. (1969). SMOG grading: A new readability formula. *Journal of Reading*, 12, 639-646.
- Meloy, J., & Gacono, C. (2000). Assessing psychopathy: Psychological testing and report writing. In C. Gacono (Ed.), *The clinical and forensic assessment of psychopathy* (pp. 231-249). Mahwah, NJ: Lawrence Erlbaum.
- Miller, H. (2004). Examining the use of the M-FAST with criminal defendants incompetent to stand trial. *International Journal of Offender Therapy and Comparative Criminology*, 48, 268-280.
- Millon, T. (1994). *Millon Clinical Multiaxial Inventory-III manual*. Minneapolis, MN: National Computer Systems.
- Morey, L. (1991). *Personality Assessment Inventory Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Otto, R., & Heilbrun, K. (2002). The practice of forensic psychology: A look toward the future in light of the past. *American Psychologist*, 57, 5-18.
- Otto, R., Poythress, N., Nicholson, R., Edens, J., Monahan, J., Bonnie, R., et al. (1998). Psychometric Properties of the MacArthur Competence Assessment Tool-Criminal Adjudication. *Psychological Assessment*, 10, 435-443.
- Overall, J., Henry, B., & Markett, J. (1972). Validity of an empirically derived phenomenological typology. *Journal of Psychiatry Research*, 9, 87-99.
- Podboy, J., & Kastl, A. (1993). The intentional misuse of standard psychological tests in complex trials. *American Journal of Forensic Psychology*, 11, 47-54.
- Rogers, R. (2003). Forensic use and abuse of psychological tests: Multiscale inventories. *Journal of Psychiatric Practice*, 9, 316-320.
- Rogers, R., Sewell, K., Ustadd, K., & Reinhardt, V. (1995). The Referral Decision Scale with mentally disordered inmates: A preliminary study of convergent and criterion-related validity. *Law & Human Behavior*, 19, 481-492.
- Steadman, H., Silver, E., Monahan, J., Appelbaum, P., Clark-Robbins, P., Mulvey, E., et al. (2000). A classification tree approach to the development of actuarial violence risk assessment tools. *Law and Human Behavior*, 24, 83-100.
- Taylor, M., & Heaton, R. (2001). Sensitivity and specificity of WAIS-III/WMS-III demographically corrected factor scores in neuropsychological assessment. *Journal of the International Neuropsychological Society*, 7, 867-874.
- Teplin, L., & Schwartz, J. (1989). Screening for severe mental disorder in jails: The development of the Referral Decision Scale. *Law and Human Behavior*, 13, 1-18.
- Webster, C., Douglas, K., Eaves, S., & Hart, S. (1997). *The HCR-20: Assessing the risk for violence*. Vancouver, British Columbia, Canada: Simon Fraser University.
- Webster, C., Harris, G., Rice, M., Cormier, C., & Quinsey, V. (1994). *Violence prediction scheme: Assessing dangerousness in high risk men*. Toronto, Ontario, Canada: University of Toronto.
- Wechsler, D. (1997). *WAIS-III: Administration and scoring manual*. San Antonio, TX: The Psychological Corporation.

- Wilson, J., Taylor, P., & Robertson, G. (1985). The validity of the SCL-90 in a sample of British men remanded to prison for psychiatric reports. *British Journal of Psychiatry, 147*, 400-403.
- Wright, P., & Friesen, W. (1985). The Carlson Psychological Survey with adolescents: Norms and scales reliabilities. *Canadian Journal of Behavioral Science, 17*, 389-399.
- Wright, P., Nussbaum, D., Lynett, E., & Buis, T. (1997). Forensic MMPI-2 profiles: Normative limitations impose interpretive restrictions with both males and females. *American Journal of Forensic Psychology, 15*, 19-37.