Screening and Assessment of Co-Occurring Disorders in the Justice System

SAMHSA
Substance Abuse and Mental Health

Instruments for Screening and Assessing Co-occurring Disorders

Screening and assessment of CODs in the justice system should incorporate use of standardized instruments that have been validated with offender populations. Use of standardized instruments will enhance the consistency of information gathered during this process and will promote a shared understanding of important domains to be reviewed in addressing CODs. Standardized instruments that yield summary scores and scores across different domains provide a common vocabulary for staff to communicate needs for treatment, supervision, and monitoring (Fletcher et al., 2009; Taxman, Cropsey et al., 2007) across different justice settings, such as courts, probation, and reentry from custody. However, many criminal justice programs do not administer standardized instruments (Cropsey et al., 2007; Friedmann et al., 2007) and instead use improvised screening and assessment techniques that have questionable validity and that may lead to poor outcomes among offenders who have CODs.

Given the absence of specialized screening instruments that address the multiple relevant components of CODs, several instruments (e.g., mental health, substance use, trauma/PTSD, motivation) are often combined to provide a comprehensive screening. These screening instruments are sometimes included in a battery to provide focused information regarding acute mental health and substance use needs and suitability for placement in various settings. Screening instruments for CODs should be administered concurrently with drug testing and examination of collateral information.

Key Issues in Selecting Screening and Assessment Instruments

There are several key issues in selecting screening and assessment instruments related to CODs:

- **Reliability.** The reliability of a screening instrument refers to the ability to obtain similar scores after readministering the same instrument over time or after administering the instrument by different people. Reliability can be difficult to achieve when screening justice-involved individuals who have CODs due to the changing symptom picture that may be affected by recent alcohol or other drug use, withdrawal from substances, use of psychotropic medications, or intentional malingering or dissimulation. Screening may need to be readministered if there are concerns about the accuracy of information obtained, and at minimum, interpretation of screening should include caveats about potential adverse influences on the accuracy of information.
- **Validity.** Many standardized mental health and substance use instruments are not sensitive to or specific in identifying CODs. Sensitivity refers to an ability to identify individuals with mental or substance use disorders, or both, while specificity refers to an ability to identify individuals without such disorders. Screening instruments that examine the same area (e.g., presence of a mental disorder) often have varying levels of sensitivity and specificity. These properties should be carefully examined, as the need for higher sensitivity or higher specificity will depend upon the particular

justice setting and the purpose of screening. For example, when using a mental health screen in a large prison system, it is very important to use an instrument with high sensitivity, so that mental disorders are not underidentified. In contrast, to identify substance use disorders in a large prison system for purposes of placement in residential treatment programs (e.g., Therapeutic Communities [TCs]), it is perhaps more important to use a screen with high specificity, so that inmates are not mistakenly placed in intensive treatment services.

- **Use in Criminal Justice Settings.** Not all screening and assessment instruments related to CODs have been validated for use within justice settings, although a growing number of studies have been conducted in these settings. Instruments that have not been validated in justice settings may still be used; however, caution is urged in interpreting results and research is needed to examine the accuracy of the particular instrument (e.g., in reference to similar instruments that have known psychometric properties).

Comparing Screening Instruments

Only a few studies have compared the effectiveness of mental health or substance use screening instruments in detecting the respective disorders (Peters et al., 2000; Sacks et al., 2007b). As part of the NIDA Criminal Justice–Drug Abuse Treatment Studies (CJ-DATS) network, a multisite study was conducted to identify effective screening instruments for CODs among individuals enrolled in prison-based addiction treatment (Sacks et al., 2007b). The effectiveness of the Global Appraisal of Individual Needs–Short Screener (GAIN-SS), the Mental Health Screening Form-III (MHSF-III), and the Mini International Neuropsychiatric Interview–Modified (MINI-M) were compared by examining results from the SCID, a comprehensive diagnostic interview, which served as the criterion measure. The MHSF-III and the GAIN-SS had somewhat higher overall accuracy than the MINI and had higher sensitivity than the MINI in detecting mental disorders (Sacks et al., 2007b). However, each of the mental health screens performed adequately in detecting severe mental disorders (i.e., bipolar disorder, major depressive disorder, and schizophrenia). These mental health-screening instruments were found to have somewhat higher overall accuracy among male offenders.

One study examined the effectiveness of substance use screening instruments among prisoners (Peters et al., 2000). Three instruments were found to be the most effective in identifying individuals with substance use disorders, as determined by the SCID diagnostic interview: the Simple Screening Instrument (SSI), the Texas Christian University Drug Dependence Screen V (TCUDS V), and a combined measure that consisted of the Alcohol Dependence Scale (ADS) and Addiction Severity Index (ASI)–Drug Use section. These instruments outperformed several other substance use screens, including the Michigan Alcoholism Screening Test (MAST)–Short version, the ASI–Alcohol Use section, the Drug Abuse Screening Test (DAST-20), and the Substance Abuse Subtle Screening Inventory (SASSI-2) on key measures of positive predictive value, sensitivity, and overall accuracy.

Subsequent sections describe a range of available mental health and substance screening instruments, as well as those examining both mental and substance use disorders.

Recommended Screening Instruments

A set of recommended screening instruments in the justice system is provided below and in Figure 8:

- Recommended screening instruments for mental disorders
  - » Brief Jail Mental Health Screen (BJMHS)

Mental Disorders	Substance Use Disorders	Co-occurring Disorders	Motivation & Readiness	Trauma History & PTSD	Suicide Risk
Brief Jail Mental Health Screen (BJMHS) (or) Correctional Mental Health Screen (CMHS-F/ CMHS-M) (or) Mental Health Screening Form-III (MHSF-III)	Brief Texas Christian University Drug Screen-V (TCUDS V)* (or) Simple Screening Instrument (SSI)* (or) Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST) Extended TCU Drug Screen V (TCUDS V)* and Alcohol Use Disorders Identification Test (AUDIT)* (or) Simple Screening Instrument (SSI)* and Alcohol Use Disorders Identification Test (AUDIT)*	Mini International Neuropsychiatric Interview-Screen (MINI-Screen) (or) Brief Jail Mental Health Screen (BJMHS)* and TCU Drug Screen V (TCUDS V)* (or) Correctional Mental Health Screen* (CMHS-F/ CMHS-M) and TCU Drug Screen V (TCUDS V)*	Texas Christian University Motivation Form (TCU-MotForm)* (or) University of Rhode Island Change Assessment Scale-M (URICA-M)*	Trauma History Screen (THS)* (or) Life Stressor-Checklist (LSC-R)* (or) Life Events Checklist for DSM-5* (and) Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5)*	Interpersonal Needs Questionnaire (INQ) and Acquired Capability Suicide Scale (ACSS)* (or) Beck Scale for Suicide Ideation (BSS) (or) Adult Suicidal Ideation Questionnaire (ASIQ)

* Instrument available at no cost

Figure 8. Recommended Screening Instruments

  - Correctional Mental Health Screen (CMHS-F/ CMHS-M)
  - Mental Health Screening Form-III (MHSF-III)
- Recommended screening instruments for substance use disorders
  - Texas Christian University Drug Screen V (TCUDS V) (Note: To conduct a screening that includes more detail about alcohol use, the AUDIT can be combined with the TCUDS V or the SSI instrument.)
  - Simple Screening Instrument (SSI)
  - Alcohol, Smoking, and Substance Involvement Screening Test (ASSIST)
  - TCU Drug Screen V (TCUDS V)
  - Alcohol Use Disorders Identification Test (AUDIT)*
  - Simple Screening Instrument (SSI)
  - Alcohol Use Disorders Identification Test (AUDIT)
- Recommended screening instruments for co-occurring disorders
  - Mini International Neuropsychiatric Interview-Screen (MINI-Screen)
  - Brief Jail Mental Health Screen (BJMHS) and TCU Drug Screen V (TCUDS V)
  - Correctional Mental Health Form (CMHS-F/CMHS-M) and TCU Drug Screen V (TCUDS V)
- Recommended screening instruments for motivation and readiness
  - Texas Christian University Motivation Form (TCU MOTForm)
  - University of Rhode Island Change Assessment Scale-M (URICA-M)
- Recommended screening instruments for trauma history and PTSD
  - The Trauma History Screen (THS), or
  - Life Stressor-Checklist (LSC-R), or
  - Life Events Checklist for DSM-5 (LEC-5), and
  - Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5)
- Recommended screening instruments for suicide risk
  - Interpersonal Needs Questionnaire (INQ), combined with the Acquired Capability Suicide Scale (ACSS)
  - Beck Scale for Suicide Ideation (BSS)
  - Adult Suicidal Ideation Questionnaire (ASIQ)

Specific instruments are recommended for screening of mental disorders, substance use disorders, co-occurring mental and substance use disorders, motivation and readiness for treatment, trauma/PTSD, and suicide risk. These screening instruments can generally be administered by nonclinicians and without extensive specialized training, although staff need to be knowledgeable about how to refer offenders who are positively identified by screens to appropriate services. Recommendations are based on a critical review of the research literature examining each area of screening. A comprehensive review of screening instruments in each of these areas is provided in subsequent sections and includes a discussion of positive features, concerns, and availability and pricing. In addition to the areas identified in Figure 8, screening of CODs in the justice system should also include examination of criminal risk. A wide variety of criminal risk screening and assessment instruments are available (Desmarais & Singh, 2013), although it is beyond the scope of this monograph to review these instruments.

As per the recommendations in Figure 8 to conduct a comprehensive screening that includes more detail about alcohol use, the AUDIT can be combined with the TCUDS V or the SSI instrument. When screening for trauma/PTSD, the THS, the LSC-R, and the LEC-5 instruments provide checklists for examining traumatic life events, and it is recommended that one of these instruments be used in combination with the PCL-5 screen, which identifies symptoms related to trauma/PTSD. Use of two separate

screening instruments to examine mental disorders and substance use disorders would require approximately 10–25 minutes to administer and score. Providing additional screening for trauma/PTSD, suicide risk, and motivation would increase the total amount of time required to approximately 25–35 minutes. Each of the recommended screening instruments in Figure 8 can be administered as repeated measures to examine changes over time. This information can be very useful in identifying the need for changes to treatment/case plans, the level of treatment and supervision services, and for further assessment.

Issues in Conducting Assessment and Diagnosis

As described previously, assessment of CODs is usually conducted after completing an initial screening and following referral to treatment services. If symptoms of both mental and substance use disorders are detected during screening, the assessment should examine the potential interactive effects of these disorders. Criminal risk factors should also be assessed, particularly the set of "criminogenic needs" or "dynamic" risk factors that can change over time and that should be the targets of justice-system interventions. Assessment provides the basis for developing an individualized treatment/case plan, and depending upon the setting, a community reentry plan. Key elements of CODs assessment include examination of skill deficits, the need for psychotropic medications, and types of treatment and ancillary services that are needed. Sufficient time should be allowed prior to assessment to ensure that an individual is detoxified and to ascertain whether any mental health symptoms exhibited are related to recent substance use (e.g., withdrawal symptoms). Standardized assessment methods should be implemented at early stages of involvement in the justice system and at key transition points during subsequent involvement in the justice system. Use of formal assessment and diagnostic instruments should be supplemented by information from collateral sources (e.g., from family members) and from archival records (e.g., criminal history).

An important component of assessment in the justice system is formal diagnoses of mental and substance use disorders. Among individuals who have CODs, this process often involves differentiating between several types of disorders (e.g., depression, anxiety, PTSD, borderline disorders) that share common symptoms and examining the potential effects of substance use on symptoms of various mental disorders. In addition to providing descriptive and prognostic information, diagnostic classification (e.g., through use of the DSM-IV-TR/DSM-5; APA, 2000, 2013) with justice-involved individuals who have CODs assists in identifying key areas to be addressed during psychosocial assessment and in developing an individualized treatment/case plan (ASAM, 2013; Stallvik, & Nordahl, 2014). Important revisions have been made to the DSM-5 criteria for both mental and substance use disorders, and these should be carefully reviewed before providing diagnoses.

A range of diagnostic instruments are available to examine symptoms of mental and substance use disorders within the DSM-5 classification framework. Instruments may be fully structured (e.g., AUDADIS-IV), thereby requiring minimal training to administer, or may be semistructured (e.g., SCID-IV), requiring training and application of clinical judgment. For a detailed review of available diagnostic instruments for examining CODs in the justice system, refer to the section "Assessment and Diagnostic Instruments for Co-occurring Mental and Substance Use Disorders."

The following considerations should be reviewed in selecting and administering diagnostic instruments:

- Structured interview instruments (e.g., SCID-IV; AUDADIS-IV) are useful in providing reliable and accurate diagnosis of CODs, although these instruments often require considerable time to administer and may not be practical in all justice settings

- Diagnostic instruments should have good interrater reliability and validity
- Diagnosis should be based on observation of mental health and substance use symptoms over time, and diagnostic interviews should be supplemented by review of collateral sources of information and by drug testing, whenever feasible
- Diagnoses of individuals with CODs should be reviewed periodically, given that key symptoms often change over time (e.g., following periods of prolonged abstinence)

Recommended Instruments for Assessment and Diagnosis of Co-occurring Disorders

Few instruments have been validated for use in assessing individuals with CODs. Moreover, few studies have attempted to validate different types of assessment instruments in criminal justice settings. Given the heterogeneity of symptoms presented by individuals with CODs, it is unlikely that a single instrument will be sufficient to assess the full range of co-occurring problems or to distinguish individuals who have CODs from those who have either a mental or a substance use disorder. Therefore, when identifying CODs in the justice system, it is important to combine different types of screening and assessment instruments to gain a comprehensive picture of psychosocial functioning and potential treatment and supervision needs (Steadman et al., 2013).

An integrated approach for assessing CODs in the justice system should include a comprehensive review of mental and substance use disorders, an examination of criminal justice history and status, and assessment of criminal risk (Steadman et al., 2013; Kubiak et al., 2011). Assessment should also review the interactive effects of mental and substance use disorders. Several previously described screening instruments may be used as part of an assessment battery to examine specialized areas (e.g., trauma history/PTSD) related to CODs. The Suicide Risk Decision Tree should be administered if suicide risk is indicated by one of the screening tools described in Figure 7. The PSS-I or PDS should also be administered if an individual endorses "high risk" on screens used to identify trauma/PTSD. These instruments can assist in differential diagnosis of PTSD and other mental disorders.

Recommendations assessment instruments are provided below and in Figure 9:

- Recommended instruments for mental disorders
  - » Personality Assessment Inventory (PAI)
- Recommended instruments for substance use disorders and treatment matching
  - » TCU Drug Screen V (TCUDS V)
  - » TCU Client Evaluation of Self and Treatment (TCU CEST)
  - » TCU Mental Trauma and PTSD Screen (TCU TRMA)
  - » TCU Physical and Mental Health Status Screen (TCU HLTH)
  - » TCU Criminal Justice Comprehensive Intake (TCU CJ CI)
- Recommended assessment and diagnostic instruments for co-occurring disorders
  - » Alcohol Use Disorders and Associated Disabilities Interview Schedule-IV (AUDADIS-IV)
  - » Mini International Neuropsychiatric Interview (MINI)
  - » Structured Clinical Interview for DSM
- Recommended assessment instruments for trauma history and PTSD
  - » The Posttraumatic Symptom Scale (PSS-I)
  - » The Posttraumatic Diagnostic Scale (PDS)
  - » Clinician Assisted PTSD Scale for DSM-5 (CAPS-5)
- Recommended assessment and diagnostic instruments for suicide risk
  - » Suicide Risk Decision Tree

These instruments are based on a critical review of the research literature examining both assessment and diagnostic instruments for use with CODs. A comprehensive review of assessment and diagnostic instruments ("Assessment and Diagnostic Instruments for Co-occurring Mental and Substance Use Disorders") is provided in subsequent sections and includes a discussion of positive features, concerns, and availability and pricing. Assessment instruments differ significantly in their coverage of areas related to mental and substance use disorders, validation for use in community and criminal justice settings, cost, scoring procedures, and training required for administration.

Assessment instruments generally require from 45–90 minutes to administer. Depending on the individual symptom presentation, administration of diagnostic instruments can require up to two hours. Selection of assessment and diagnostic instruments should consider the level of staff training, certification, and expertise required.

Screening Instruments for Substance Use Disorders

A wide range of substance use screening instruments are available, including both public domain and proprietary products. These instruments vary considerably in their effectiveness, cost, and ease of administration and scoring (Hiller et al., 2011). As with other screening instruments, substance use screens are somewhat vulnerable to manipulation by those seeking to conceal substance use problems, and concurrent use of drug testing is recommended to generate the most accurate screening information

Mental Disorders	Substance Use Disorders and Treatment Matching	Co-occurring Disorders	Trauma History and PTSD	Suicide Risk
Personality Assessment Inventory (PAI)	TCU Drug Screen V (TCUDS V)*, TCU Client Evaluation of Self and Treatment (TCU CEST)*, TCU Mental Trauma and PTSD Screen (TCU TRMA)*, and TCU Physical and Mental Health Status Screen (TCU HLTH)* (and/or) TCU Criminal Justice Comprehensive Intake (TCU CJ CI)*	Alcohol Use Disorders and Associated Disabilities Interview (AUDADIS-IV)* (or) Mini International Neuropsychiatric Interview (MINI) (or) Structured Clinical Interview for DSM (SCID)	Posttraumatic Symptom Scale (PSS-I)* (or) Posttraumatic Diagnostic Scale (PDS) (or) Clinician Assisted PTSD Scale for DSM-5 (CAPS-5)*	Suicide Risk Decision Tree (SRDT)*

*Instrument available at no cost

Figure 9. Recommended Assessment Instruments

(Richards & Pai, 2003). A range of substance use screening instruments are reviewed in this section that can assist in detecting co-occurring disorders (CODs), with information provided about positive features and concerns related to each instrument.

Changes to the DSM-5 Diagnostic Classification System

Several substance use disorders are described in the section of the DSM-5 (APA, 2013) entitled "Substance-Related and Addictive Disorders." Substance use and substance dependence are no longer considered separate disorders as they were in DSM-IV, and have been combined into a single disorder ("substance use disorder") that measures severity of symptoms on a continuous scale from mild to severe. The new DSM-5 resolves a problem with the DSM-IV approach, which classified "substance abuse" as a milder form of "substance dependence" when in fact the symptoms of substance misuse can be quite severe in clinical practice. On the other hand, "substance dependence" can imply that the individual is psychologically addicted to the substance when in fact the individual may be physically dependent on the substance, which is a normal physiological response to certain drugs.

Major highlighted changes to the DSM-5 classification system for substance use disorders are as follows:

- There are a total of 11 symptoms of substance use disorders that combine elements of DSM-IV "abuse" and "dependence" diagnostic criteria
- "Mild" substance use disorder requires endorsement of 2–3 symptoms out of a total of 11 symptoms
- "Moderate" substance use disorders require the presence of 4-5 symptoms, while "severe" disorders require 6 or more symptoms
- Changes from the DSM-IV classification of substance "abuse" and "dependence" disorders to the DSM-5 classification of "mild," "moderate," and "severe" substance use disorders have not apparently affected the prevalence of alcohol or drug use diagnoses in offender populations (Kopak, Proctor, & Hoffman, 2014)
- Gambling disorder is an addictive disorder resembling substance use disorders from the biopsychosocial perspective
- Caffeine disorder is no longer considered an addictive disorder

Screening Instruments

Alcohol Dependence Scale (ADS)

The ADS (Skinner & Horn, 1984) is a widely used 25-item instrument developed to screen for symptoms of alcohol use disorders. This measure assesses withdrawal symptoms, increased alcohol tolerance, awareness of compulsive and excessive drinking, salience of drink-seeking behaviors, and impaired control over drinking. The instrument was developed through factor analysis of the original 147-item Alcohol Use Inventory (AUI) and is published by the Addiction Research Foundation. Questions on the ADS are specific to the last 12 months and can be given as a clinical interview or self-report assessment (Chantarujikapong, Smith, & Fox, 1997). A cut-off score of ≥ 8 has been used in clinical samples to identify those with alcohol use diagnoses (Chantarujikapong et al., 1997; Ross, Gavin, & Skinner, 1990). Only 9 of the 25 ADS items may be needed to make a reliable classification in high-risk alcohol drinkers, and ADS items addressing excessive drinking are the most useful in making this classification (Kahler, Strong, Stuart, Moore, & Ramsey, 2003; Kahler, Strong, Hayaki, Ramsey, & Brown, 2003).

Positive Features

- The ADS is brief, inexpensive, easily scored, and does not require specialized training to administer
- The ADS has been found to perform adequately in community settings (Ross et al., 1990)

- The ADS is unidimensional, as intended, and has good internal consistency (alpha = .90; Kahler, Strong, Stuart et al., 2003)
- ADS scores are significantly correlated with objective measures of alcohol use severity among incarcerated men (Hodgins & Lightfoot, 1989)
- The ADS is most effective in detecting moderate to severe levels of alcohol use (Chantarujikapong et al., 1997)
- The ADS in combination with the Addiction Severity Index (ASI)–Drug Use section was one of three screening instruments found to be the most effective in identifying substance use among prisoners (Peters & Greenbaum, 1996)
- The ADS was the most accurate of several screening instruments in detecting alcohol disorders among justice-involved individuals (Peters et al., 2000)
- In determining substance use disorders among offenders, the ADS exhibited adequate sensitivity (74 percent, 66 percent), specificity (92 percent, 97 percent), positive predictive value (89 percent, 98 percent), and negative predictive value (80 percent, 69 percent) respectively (Peters et al., 2000)
- The ADS performed as well as the Michigan Alcoholism Screening Test (MAST) in detecting alcohol use disorders (Ross et al., 1990)
- In an addictions setting, at a cut-off score of 8 or 9, the ADS has good sensitivity (91 percent), specificity (82 percent), positive predictive value (93 percent), and negative predictive value (76 percent; Ross et al., 1990)
- A 12-item version of the ADS can reliably discriminate between levels of alcohol severity in treatment-seeking populations (Kahler, Strong, Hayaki et al., 2003)
- The ADS provides both cut-off scores that indicate the presence of an alcohol use disorder and treatment
- The ADS has been found to have test-retest reliability of .92–.98 over a 1-week period (Addiction Research Foundation, 1993; Peters et al., 2000)
- Computerized versions of the ADS are available through the Computerized Lifestyle Assessment. Miller and others (2002) report high test-retest reliability of this version (r score = .84–.93) over a 1-week period

Concerns

- The ADS does not examine quantity or frequency of recent and past alcohol use
- The ADS is limited to screening for alcohol use problems
- The superficial nature of ADS items may result in underreporting of symptoms
- Additional validation in subpopulations may be necessary (e.g., pregnant women)
- The ADS does not always exhibit substantial agreement across types of reporting (e.g., self-report, report by service/agency staff), with one study indicating only a 15 percent rate of agreement in a treatment-seeking population
- The ADS is a commercial product, although the cost is quite modest

Availability and Cost

The ADS is a copyrighted document that can be obtained from its author. The price of $15 includes a user's guide and 25 questionnaires. Additional packets of 25 questionnaires cost $6.25. Requests for the kits can be made to Harvey Skinner Ph.D., Department of Public Health Sciences, McMurrich Building, University of Toronto, Toronto, Ontario, Canada M5S 1A8. E-mail requests can be sent to harvey.skinner@utoronto.ca

The ADS can be downloaded at no cost at the following site: http://www.emcdda.europa.eu/html.cfm/index3583EN.html

Computerized versions of the ADS can be obtained by contacting the Multi-Health Systems regarding and requesting the Computerized Lifestyle Assessment: 1-800-456-3003 (U.S.); 1-800-268-6011 (Canada).

Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)

The ASSIST (World Health Organization [WHO] ASSIST Working Group, 2002) was developed for the WHO by an international group of substance use researchers to address the need for a comprehensive screening instrument in primary health care settings. The original 12-item instrument was developed through identifying psychometrically sound items from other substance use screens, based on a comprehensive review of the literature (Babor, 2002). The ASSIST measures frequency of substance use; current symptoms (i.e., in the past 3 months); and problems related to alcohol, tobacco, and other drugs. The ASSIST includes a brief introduction describing the purpose of the measure, and items are grouped by type of substance (e.g., alcohol, cannabis, opioids, stimulants, tobacco). Item 1 provides a brief screen for lifetime use of each type of substance.

The remaining items on the ASSIST examine current frequency of substance use by type of substance, and frequency of related symptoms during the past 3 months. For example, item 2 inquires about current frequency of use ("how often have you used the substance in the past 3 months?"). Subscales of the ASSIST include Specific Substance Involvement (SSI; sum of items 2–7 for each type of substance) and Total Substance Involvement (TSI; sum of items 1–8 across each type of substance). Item 8 inquires about intravenous (IV) drug use in the past 3 months. The ASSIST provides feedback to respondents indicating the level of their SSI score by severity of risk for substance use problems according to designated cut-off scores (low risk = 0–3, moderate = 4–26, high ≥ 27) and physical and mental health risks associated with these scores. The risk levels are also intended to distinguish between low, medium, and high risk. An integrated set of brief interventions provides feedback regarding health risks for each substance class.

Modifications to the instrument (ASSIST 2.0) reduced the number of items to eight, and improved the psychometric properties. The most recent version (ASSIST 3.0) provides standardized cut-off scores across different types of substances. The NIDA has modified this measure to include two parts: (1) the "NIDA Quick Screen," and (2) the "NIDA Modified ASSIST," which provides a more comprehensive assessment for individuals who surpass the cut-off score on the Quick Screen. The Quick Screen inquires only about past year use of alcohol, tobacco, and drugs. The ASSIST has been widely adapted for use in different cultures and has been translated into several languages. This instrument can be administered as an interview or by self-report.

Positive Features

- The ASSIST is available at no cost, is quite brief to administer, and includes scoring and interpretation of scores (e.g., level of treatment needs) according to risk level
- The ASSIST evaluates lifetime substance use, current substance use, severity of substance use, and risk related to IV drug use
- The ASSIST 3.0 includes weighting and recoding analyses that provide a consistent cut-off score for substance use
- The ASSIST uses an approach that is consistent with the federally funded Screening, Brief Intervention, and Referral to Treatment (SBIRT) initiative in that accompanying materials are provided to implement brief interventions and referral to treatment, based on ASSIST findings related to risk level and type of substance(s) used
- The ASSIST includes cut-off scores for differentiating between severity of use (low risk: ≤ 3; moderate risk: ≤ 26; and

high risk: ≥ 27), and is able to adequately distinguish between these risk categories across different types of substances (Humeniuk et al., 2008)

- The ASSIST 2.0 (Humeniuk et al., 2008) has been validated in several countries, using samples that are balanced across age and gender
- The ASSIST 2.0 demonstrates good overall psychometric properties (Humeniuk et al., 2008). In terms of concurrent validity, the frequency of current use for each type of substance (item 2) is highly correlated with the Addiction Severity Index (ASI; r scores range .76–.88), and the total substance involvement scores (TSI) are highly correlated with total MINI (Mini Neuropsychiatric Interview) substance use disorder diagnoses (r score =.76) and with scores on the SDS (Severity of Dependence), the RTQ (Revised Fagerstrom Tolerance Questionnaire), and the Alcohol Use Disorders Identification Test (AUDIT)
- The ASSIST scores are associated with physical and mental health problems, as well as IV drug use (Humeniuk et al., 2008)
- The ASSIST 2.0 TSI and SSI scores demonstrate adequate to good sensitivity and specificity in distinguishing between differently levels of use. Finally, the ASSIST scores showed strong correlations with the MINI diagnoses (Humeniuk et al., 2008)
- Kappa reliabilities for agreement between test administrations in the original validation study of the ASSIST 1.0 (WHO Group, 2002) were adequate (kappas range .58–90)
- The ASSIST 2.0 demonstrates good internal consistency (alphas range .77–.94) across different types of substances (Humeniuk et al., 2008)
- The single item Quick Screen from the NIDA-modified ASSIST provides good sensitivity (100 percent) and adequate specificity (74 percent) in classifying individuals with substance use disorders. These results are comparable to those obtained from the Drug Abuse Screening Test, DAST-10 (Smith, Schmidt, Allensworth-Davies & Saitz, 2010)

Concerns

- The ASSIST has not been widely studied in offender populations
- Caution should be exercised when interpreting the different ASSIST risk levels for substance use problems, as the instrument appears to more effectively distinguish between low and moderate risk than between moderate and high risk for each type of substance, as measured by SSI scores and by the Total Substance Involvement scores (TSI). Additional studies are needed to examine the ability of the ASSIST to discriminate between the different risk levels (Humeniuk et al., 2008)
- The cut-off score for alcohol risk levels (≤ 10, low risk; ≤ 26, moderate risk; ≥ 27, high risk) is different from the scores for other substances (Humeniuk et al., 2008)
- Validation results for the ASSIST may be inflated by reliance on self-report information
- Further studies of the ASSIST are needed to determine the instrument's validity by gender, culture, race/ethnicity, and language
- Further work is also needed to examine the utility of the ASSIST in providing triage to therapeutic interventions in primary care settings
- Studies have not investigated the differential effects on validity of the interview and self-report versions of the ASSIST
- The NIDA-modified ASSIST does not provide detailed risk assessment feedback, as does the original ASSIST
- A one-item screen for drug use in the past year (such as the NIDA Quick Screen) may be less accurate in determining current

substance use among men and Hispanics, relative to other groups (Smith et al., 2010)

Availability and Cost

The most recent version of the ASSIST (3.0) is available at no charge via electronic download and includes the screening tool, user's manual, patient feedback card, as well as self-help strategies for managing substance use. The instrument can be obtained at the following site: http://www.who.int/substance_abuse/activities/assist/en/index.html

The NIDA-modified ASSIST is available at no charge via electronic download at the following site, which includes detailed instructions for administration and scoring: http://www.drugabuse.gov/sites/default/files/pdf/nmassist.pdf

Alcohol Use Disorders Identification Test (AUDIT)

The AUDIT is a two-part screening instrument that was developed by the World Health Organization (WHO). The AUDIT is based on the International Classification of Disease-10 (ICD-10) criteria and is intended to identify individuals who have harmful levels of drinking in order to prevent harmful consequences. The instrument was initially developed for screening in primary health care settings and was designed for use in multiple cultures and settings to assess harmful and hazardous alcohol use in the past year. Studies indicate that the AUDIT examines three major factors: (1) alcohol consumption, (2) drinking behaviors, and (3) consequences of drinking.

The first part of the instrument (AUDIT Core) is a brief, 10-item questionnaire created to measure alcohol consumption, symptoms, and alcohol-related consequences. The second part of the instrument (AUDIT-CSI, Clinical Screening Instrument) is a supplement to the Core and assesses physiological consequences of alcohol use. The CSI consists of three sections: (1) trauma history, (2) abnormal physical exam findings, and (3) serum gamma-glutamyl transpeptidase level, which identifies harmful effects of alcohol use. Several brief forms of the AUDIT include the three-item AUDIT-C screen (Bush, Kivlahan, McDonell, Fihn & Bradley, 1998), the FAST, a four-item screening form (Hodgson, Alwyn, John, Thom & Smith, 2002), and the five-item AUDIT-5 (Kim et al., 2013).

The recommended cut-off score on the AUDIT for identifying hazardous drinking or alcohol use disorders is ≥ 8, and cut-off scores on the AUDIT-C are ≥ 4 with men and ≥ 3 with women (Babor, Higgins-Biddle, Saunders & Monteiro, 2001; Bush et al., 1998). The AUDIT can be administered as an interview or as a self-report instrument. Both computerized and paper and pencil versions of the AUDIT are available, and there do not appear to be significant differences in the accuracy of information produced by these different versions (Lieberman, 2003, 2005; Saitz et al., 2004; Chan-Pensley, 1999). Many foreign language versions of the AUDIT have been developed. Although the psychometric properties of these versions have improved over time, they are still somewhat uneven across versions of the instrument (Reinart & Allen, 2007).

Positive Features

- The AUDIT is quite brief to administer and easy to read, requiring only a seventh grade reading level
- Items were carefully selected based on factor analytic procedures (Bohn, Babor, & Kranzler, 1995)
- The AUDIT appears to have two distinct factors across adult and adolescent populations, including consequences of drinking and alcohol consumption (Carey, Carey & Chandra, 2003; Doyle, Donovan, & Kivlahan, 2007; Karno, Granholm & Lin, 2000; Maisto, Conigliaro, McNeil, Kraemer & Kelly, 2000; von der Pahlen et al., 2008; Rist, Glöckner-Rist, & Demmel, 2009; Shevlin & Smith, 2007; Shields, Guttmannova, & Caruso, 2004)
- The AUDIT has been shown to predict alcohol withdrawal syndrome (Dolman

& Hawkes, 2005; Reinert & Allen, 2007; Reoux, Malte, Kivlahan & Saxon, 2002)

- The AUDIT provides cut-off scores that indicate alcohol severity and risk level, interpretation of these cut-off scores, and treatment recommendations (Babor et al., 2001)
- The AUDIT has adequate sensitivity and specificity using the standard cut-off score of 8 (Shields & Caruso, 2003). This cut-off score is most useful in detecting alcohol use disorders, while lower cut-off scores are advisable for detecting hazardous drinking (Maisto & Saitz, 2003)
- The AUDIT is a reliable and valid indicator of problem drinking among people who have serious mental illness (Cassidy, Schmitz, & Malla, 2008; Maisto, Carey, Carey, Gordon, & Gleason, 2000; Maisto, Conigliaro et al., 2000; O'Hare, Sherrer, LaButti, & Emrick, 2004; Carey et al., 2003; Reinert & Allen, 2002) and has high sensitivity and specificity for alcohol use disorders among this population (Cassidy et al., 2008; Dawe, Seinen, & Kavanaugh, 2000; O'Hare et al., 2004; Maisto, Carey et al., 2000, Maisto, Conigliaro et al., 2000)
- The AUDIT demonstrates good convergence with the SCID among psychiatric populations (Cassidy et al., 2008; Maisto, Carey et al., 2000; Maisto, Conigliaro et al., 2000). The optimal cut-off score for the AUDIT is 10 with psychiatric populations, which provides sensitivity of 85 percent, specificity of 91 percent, positive predictive value of 65 percent, and negative predictive value of 97 percent (Cassidy et al., 2008)
- The AUDIT has generally performed well across a variety of settings and populations. The instrument's internal consistency is good, with a median alpha of .83 (alphas range .75–.97; Lima et al., 2005; Reinert & Allen, 2007; Selin, 2003; Shields et al., 2004)
- Among community samples, the AUDIT demonstrates good accuracy (kappas range .70–.89) in classifying alcohol use disorders (e.g. positive or negative AUDIT score) at a cut-off score of 8 (Dybek et al, 2006; Reinert & Allen, 2007; Rubin et al., 2006; Selin, 2003)
- The sensitivity of the AUDIT is quite high in comparison to the Michigan Alcoholism Screening Test (MAST) and the CAGE (Cherpitel, 1998). The AUDIT appears to be one of the most sensitive instruments in detecting current alcohol use disorders across different populations and is quite effective in identifying low-level hazardous drinking
- The AUDIT has good sensitivity (81–85 percent), specificity (86–89 percent) and adequate positive predictive value (65 percent; Skipsey, Burleson, & Kranzler, 1997) for alcohol use disorders among substance-involved treatment populations (Pal, Jena, & Yadav, 2004; Skipsey et al., 1997)
- The AUDIT is more accurate than the CAGE or the Short Michigan Alcoholism Screening Test (SMAST-G) in identifying problematic alcohol use among the elderly (Moore, Seeman, Morgenstern, Beck & Reuben, 2002) and has good psychometric properties with middle-aged men and elderly psychiatric patients (Philpot et al., 2003; Tuunanen, Aalto, & Seppä, 2007)
- The AUDIT is equally reliable across gender, ethnic/racial, and age groups (Cherpitel, 1997; Kokotailo et al., 2004; McCloud, Barnaby, Omu, Drummond, & Aboud, 2004; Selin, 2003; Shields & Caruso, 2003; Steinbauer, Cantor, Holzer & Volk, 1998; Volk, Steinbauer, Cantor, & Holzer, 1997)
- The AUDIT has good test-retest reliability (.84–.95) over a 30-day interval (Dybek et al., 2006; Kim, Gulick, Nam & Kim, 2008; Reinert & Allen, 2007; Selin, 2003)
- The AUDIT has good psychometric properties (particularly sensitivity and specificity) across a variety of ethnic groups, including White non-Hispanic,

Hispanic, Asian, and African American men and women (Adewuya, 2005; Cherpitel, 1998; Meneses-Gaya et al., 2010; DeSilva, Jayawardana, & Pathmeswaran, 2008; Gomez et al., 2006; Giang et al., 2005; Wu et al., 2008), and is effective in identifying risky drinking and alcohol use disorders among a variety of populations (Cassidy et al., 2008; Caviness et al., 2009; DeSilva et al., 2008; Doyle et al., 2007; Meneses-Gaya et al., 2010; Tuunanen, et al, 2007)

- The AUDIT has good sensitivity and adequate specificity in identifying risky drinking and alcohol use disorders among college students (Kokotailo et al., 2004)
- Non-English versions of the AUDIT provide adequate internal consistency (Reinhert & Allen, 2007). Test-retest reliability of these versions are also acceptable (kappas range .69–.86; Dybek et al., 2006; Selin, 2003)
- The AUDIT-C demonstrates good sensitivity and specificity (81–95 percent and 73–91 percent, respectively) for identifying harmful drinking patterns and current alcohol use disorders at varying cut-off scores (ranging 2–7) across groups that differ by gender, population, and culture (Bradley et al.,2007; Bradley et al., 2003; Caviness et al., 2009; Dawson, Grant, Stinson & Zhou, 2005; Frank et al., 2008; Gual, Segura, Contel, Heather, & Colom, 2002; Seale et al., 2006)
- The AUDIT-C demonstrates good internal consistency in both clinical and college samples (.74 and .81 respectively; Shields et al., 2004) and high test-retest reliability (r score = .98; Bergman and Kallman, 2002)
- The FAST has been validated in several settings and demonstrates good psychometric properties (Hodgson et al., 2002). The FAST is correlated with other well-validated screening measures of alcohol use disorders, including the AUDIT, PAT (Paddington Alcohol Test), and the CAGE. The FAST has good sensitivity (91 percent) and specificity (93 percent) in detecting alcohol use disorders and demonstrates better psychometric properties than the CAGE and PAT (Hodgson et al., 2002)
- Among adolescents, the AUDIT has greater sensitivity than the CAGE in detecting alcohol use disorders of varying severity (Knight, Sherritt, Harris, Gates, & Chang, 2003) and has been shown to have good concurrent and criterion validity (Kelly, Donovan, Kinnane, & Taylor, 2002; Knight et al., 2003) and reliability (Kelly et al., 2002). No gender differences were found in using the AUDIT among adolescent inpatients (Kelly et al., 2002). At a cut-off score of 2 for identifying problematic alcohol use among adolescents, the AUDIT's sensitivity was 88 percent and the specificity was 81 percent (Knight et al., 2003)

Concerns

- The AUDIT does not examine substance use problems occurring prior to the last year, and is more effective in detecting current rather than previous alcohol problems (McCann, Simpson, Ries, & Roy-Byrne, 2000)
- There is considerable variability in the AUDIT-C cut-off scores by gender, culture, and population (Seale et al., 2006; Bradley et al., 2003; Dawson, Grant & Stinson, 2005; Dawson, Grant, Stinson, & Zhou, 2005; Gual et al., 2002)
- The instrument has only moderate specificity (74 percent for the "Core" and 40 percent for the "Clinical" component [Bohn et al., 1995])
- There has been little research examining the temporal stability of the AUDIT in different populations
- Within a DUI sample, the AUDIT was found to be less effective in detecting substance use disorders than the MAST (Conley, 2001)

- The AUDIT has lower reliability in alcohol drinkers with low levels of consumption
- The AUDIT may be more effective in identifying needs for assessment and treatment for justice-involved individuals when conducted several weeks after entry to prison (Maggia et al., 2004), as shown by the weak agreement in classification between initial screening and later screening (kappa = .27)
- The AUDIT-CSI is somewhat invasive and must be conducted by a trained clinician
- The AUDIT-C may be better at identifying alcohol use disorders in women than men (Dawson, Grant, Stinson, & Zhou, 2005)
- The AUDIT and the AUDIT-C are less sensitive and more specific with females (Reinert & Allen, 2002; Bradley et al., 2003) and are generally more effective screens for alcohol use disorders among women (Dawson, Grant, Stinson, & Zhou, 2005)
- Some have recommended that cut-off scores should be lowered when the AUDIT and AUDIT-C are used with women, and these scores have varied across female samples (Bradley et al., 2007; Bradley et al., 2003; Chung, Colby, Barnett, & Monti, 2002; Gache et al., 2005; Gual et al., 2002; Neumann et al., 2004), although there is little research to validate the use of specific cut-off scores for this purpose
- AUDIT-C item 3 may contribute to the sensitivity and specificity differences (Bradley et al., 2003) among female respondents
- The AUDIT has not been found to be highly accurate with the elderly in different populations (Philpot et al., 2003; Moore, Beck, Babor, Hays, & Reuben, 2002; Reinert & Allen, 2002) and has low sensitivity but good specificity with this population (O'Connell et al., 2004)
- The AUDIT-C may have lower sensitivity (43-46 percent) in primary health care settings (Seale et al., 2006)
- The AUDIT may perform more poorly among African Americans in comparison to Whites (Cherpitel & Bazargan, 2003)
- The AUDIT does not perform consistently well across all domains in identifying alcohol use disorders among adolescents and may need items that are better tailored for this age group (Chung et al., 2002)
- More research is needed to determine acceptable cut-off scores for the AUDIT among non-English speaking populations and in international settings (Cherpitel, Ye, Moskalewicz & Swiatkiewicz, 2005; Pal et al., 2004; Rumpf, Hapke, Meyer & John, 2002; Tsai, Tsai, Chen & Liu, 2005)

Availability and Cost

The AUDIT: Guidelines for Use in Primary Care Settings-Second Edition is available free of charge from the WHO at the following site: http://whqlibdoc.who.int/hq/2001/WHO_MSD_MSB_01.6a.pdf

The interview and self-report versions of the AUDIT, with scoring rules, are available at the following site: http://www.drugabuse.gov/sites/default/files/files/AUDIT.pdf

Comprehensive guidelines for use of the instrument are available from the WHO at the following site: http://whqlibdoc.who.int/hq/2001/WHO_MSD_MSB_01.6a.pdf

The AUDIT-C is available at no cost and is available with information describing scoring and interpretation at the following site: http://www.integration.samhsa.gov/images/res/tool_auditc.pdf

CAGE

The CAGE is a brief four-item screen to identify alcohol use problems (Mayfield, McCleod, & Hall, 1974). The CAGE is among the most widely used brief alcohol screening instruments with adults (Bastiaens, Riccardi, & Sakhrani, 2002). The four questions corresponding to the acronym CAGE consist of the following: (1) Have you felt you ought to Cut down on your drinking?, (2) Have

people Annoyed you by criticizing your drinking?, (3) Have you ever felt bad or Guilty about your drinking?, and (4) Have you had a drink first thing in the morning to steady your nerves or to get rid of a hangover (Eye-opener)? A total score is obtained to reflect the level of alcohol use severity.

Although the CAGE reviews lifetime alcohol problems, the National Institute on Alcohol Abuse and Alcoholism (NIAAA) has developed a version of the CAGE that examines problems during the past year. This past year version of the CAGE is more specific but less sensitive than the traditional CAGE (Bradley, Kivlahan, Bush, McDonnell, & Fihn, 2001). The CAGE can be administered via self-report or interview, and similar outcomes are obtained using both approaches (Aertgeerts, Buntix, Fevery, & Ansoms, 2000). A computerized version of the CAGE/CAGE-Adapted to Include Drugs (CAGE-AID; see "Positive Features" below) is also available, and this method has yielded higher rates of illegal drug use and substance use problems than administration through interview (Turner et al., 2005). There are alternative versions to the CAGE that include other items from the AUDIT and the MAST, such as the Augmented CAGE (Bradley, Bush, McDonnell, Malone, & Fihn, 1998), the "5-shot" (Seppä, Lepistö, Sillanaukee 1998) and the Leubeck Alcohol Dependence and Abuse Screening Test (LAST) Questionnaire (Rumpf, Hapke, Hill, & John, 1997).

The CAGE-AID is a four-item instrument that screens for both alcohol and other drug use disorders (Brown & Rounds, 1995). More in depth screens are also available that combine the CAGE-AID with other drug use questions (e.g., TICS or CRAFFT instruments). The recommended cut-off score for identifying possible alcohol problems in the CAGE is ≥ 2 positive responses (Cherpitel, 1997), in the 5-shot is ≥ 3 positive responses (Seppä et al., 1998), in the Augmented CAGE is ≥ 2 positive responses (Bradley, Bush et al., 1998), and in the LAST is ≥ 2 (Rumpf et al., 1997). The recommended cut-off score in identifying probable alcohol or drug problems with the CAGE-AID is ≥ 2 positive responses (Brown & Rounds, 1995).

Positive Features

- The CAGE does not require specific training and can be administered by a nonclinician
- The CAGE is quite brief to administer
- At a cut-off score of 1 or 2, the CAGE exhibits good sensitivity (82–91 percent), specificity (83–94 percent), and positive predictive value (74–85 percent) in classifying alcohol use disorders among patients who have schizophrenia (Dervaux et al., 2006)
- The CAGE has moderately good sensitivity (74 percent) and very good specificity (97 percent) in diagnosing substance use disorders among individuals with schizophrenia (McHugo, Paskus, & Drake, 1993) and generally has been shown to have good sensitivity and specificity among clinical populations (Bastiaens et al., 2002)
- Among inpatient populations, the CAGE exhibits adequate sensitivity (87 percent) and specificity (77 percent) at a cut-off score of 2 for alcohol use disorders
- The CAGE has higher sensitivity in diagnosing alcohol use disorders in inpatient populations than in other settings (Aertgeerts, Buntinx, & Kester, 2004)
- In a primary care population, the CAGE exhibits adequate sensitivity (85 percent) and specificity (78 percent) at a cut-off score of 1 for alcohol use disorders (Aertgeerts et al., 2004)
- The CAGE exhibits adequate sensitivity (62–89 percent) and specificity (79–93 percent) among different racial/ethnic groups at a cut-off score of 2 (Buchbaum, Buchanan, Centor, Schnoll, & Lawton, 1991; Dhalla & Kopec, 2007; Saremi et al., 2001; Saitz, Lepore, Sullivan, Amaro & Samet, 1999)
- Diagnostic agreement between written and interview versions of the CAGE is quite good (k = .83; Aertgeerts et al., 2000), as

is agreement between computerized and in-person interviews (.77; Bernadt, Daniels, Blizard & Murray, 1989)

- Internal consistency of the CAGE across clinical and nonclinical samples averages .74 (Shields & Caruso, 2004)
- The CAGE is highly correlated with other validated measures of alcohol use disorders, such as the SMAST (Hays & Merz, 1995), and the CAGE-AID is highly correlated with the AUDIT (Leonardson et al., 2005), supporting the convergent validity of these instruments
- The test-retest reliability of the CAGE was found to be .80 among psychiatric outpatients, and .95 in a community sample (Teitelbaum & Carey, 2000)
- The CAGE more effectively classifies college students than the SASSI-3 (Clements, 2002). The CAGE has also been found to effectively distinguish between adolescents who have alcohol use disorders and those who do not have these disorders (Hays & Ellickson, 2001)
- The CAGE-AID has greater sensitivity and lower specificity for substance use disorders in comparison to the CAGE. The CAGE-AID has greater sensitivity than the CAGE across gender, income, education, and different types of substance use disorders (Brown & Rounds, 1995)
- The CAGE-AID shows high internal consistency (r score= .92; Leonardson et al., 2005)

Concerns

- The CAGE does not examine quantity or frequency of recent and past substance use and examines a narrow range of diagnostic symptoms related to alcohol use disorders
- The CAGE has not been widely validated for use in justice settings
- The CAGE may have lower test-retest reliability among psychiatric patients than in other populations (r score = .67; Dyson et al., 1998)
- The reliability of the CAGE ranges greatly (.52–.90) across different samples (Shields & Coruso, 2004)
- Interrater reliability of the CAGE for diagnosis of substance use disorders is quite low (kappa = .15; Indran, 1995)
- The CAGE does not effectively discriminate between heavy and non-heavy drinking in the general population (Bisson, Nadeau, & Demers, 1999). Due to the focus on lifetime problems, the CAGE does not differentiate between people with chronic alcohol problems and those who have not experienced problems in many years (Bradley et al., 2001)
- Within general population samples, no CAGE cut-off score provides concurrently high specificity, sensitivity, and positive predictive value (Bisson et al., 1999)
- The CAGE sometimes provides low sensitivity in classifying alcohol use disorders (Maisto, & Saitz, 2003), and there is wide variability in the instrument's sensitivity (43–94 percent)
- Higher CAGE cut-off scores provide better specificity and sensitivity in primary care settings than in other settings (Aertgeerts et al., 2004)
- The CAGE is more accurate in classifying males than females (McHugo et al., 1993). The instrument underestimates alcohol problems among females (Bisson et al., 1999; Cherpitel, 2002; Matano, Wanat, Westrup, Koopman & Whitsell, 2002; Moore, Beck et al., 2002). The CAGE also has lower sensitivity among White females than African American females (Bradley, Boyd-Wickizer, Powell, & Burman, 1998)
- The CAGE has higher sensitivity among African Americans than Whites (Cherpitel 2002)
- Translation and cultural differences may affect responses on the CAGE (Steinbauer et al., 1998)
- The CAGE has low sensitivity among elderly psychiatric samples (O'Connell et al., 2004)

- The CAGE is not recommended for use with adolescents (Hays & Ellickson, 2001; Knight et al., 2003) and has performed poorly in college samples (Aertgeerts et al., 2000; Bisson et al., 1999)
- Several alternate versions (LAST, 5-shot, Augmented CAGE) have better psychometric properties than the CAGE in detecting alcohol use problems and disorders (Bradley, Bush et al., 1998; Rumpf et al., 1997; Seppä et al., 1998)

Availability and Cost

The CAGE is available free of charge, and the instrument and scoring information can be found at either of the following sites:

- http://bit.ly/CAGE_inst
- http://www.projectcork.org/clinical_tools/html/CAGE.html

The CAGE can also be obtained in the document: Ewing, J. A. (1984). Detecting alcoholism: the CAGE questionnaire. *Journal of the American Medical Association, 252* (14), 1905–1907.

The Dartmouth Assessment of Lifestyle Instrument (DALI)

The DALI is an 18-item, interview-administered scale that examines lifetime alcohol, cannabis, and cocaine use disorders among people with severe mental illness. The DALI is a composite of several different instruments and includes 3 items from the Life-Style Risk Assessment Interview and the remaining 15 items from the Reasons for Drug Use Screening Test, the TWEAK, the CAGE, the Drug Abuse Screening Test (DAST), and the ASI. The DALI contains two scales that assess risk for alcohol use disorders and drug use disorders. It is designed for people who have more severe psychopathology (Rosenberg et al., 1998). This instrument has not been studied extensively among broad sets of clinical populations. Information about recommended cut-off scores can be obtained from the authors, as described in the following section regarding availability and cost.

Positive Features

- The DALI requires approximately 6 minutes to administer and is easy to score
- The instrument has good specificity (80 percent) and sensitivity (100 percent) in identifying substance use among people with mental disorders (Rosenberg et al., 1998)
- The DALI alcohol scale has good specificity (98 percent) and overall accuracy of 73 percent in diagnosing alcohol use disorders. The DALI drug scale has good specificity (97 percent) and average sensitivity (50 percent), with overall accuracy of 83 percent in diagnosing drug use disorders among psychiatric inpatients (Ford, 2003)
- The DALI may be good at minimizing "false positive" classifications (Ford, 2003)
- Interrater reliability ranges .86–.98 (Rosenberg et al., 1998). The DALI has been shown to have test-retest reliability of .90 (Rosenberg et al., 1998)

Concerns

- The DALI was developed and validated on newly admitted psychiatric inpatients in a predominantly White and rural population
- Future research is needed to validate its use in ethnically and culturally diverse populations, and in justice and substance use treatment settings
- The instrument only examines alcohol, cannabis, and cocaine use disorders
- The DALI alcohol screen may have low specificity among psychiatric inpatients (Ford, 2003)

Availability and Cost

The DALI, scoring instructions, cut-off scores, and reference materials can be obtained at no cost from the University of Washington Alcohol and Drug Abuse Library website: http://bit.ly/DALI_inst

The instrument and scoring instructions can also be obtained at the following site: http://www.dhs.state.mn.us/dhs16_141793.pdf

Drug Abuse Screening Test (DAST)

The DAST (Skinner, 1982) is a brief screening instrument that examines symptoms of substance use disorders. Several versions of the DAST are available, including the original DAST-28, DAST-20, DAST-10, and DAST for Adolescents (DAST-A). The DAST reviews drug and alcohol problems occurring in the past 12 months. Items from the DAST were developed to align with those developed for the Michigan Alcoholism Screening Test (MAST). The recommended cut-off score for identifying drug use disorders with the DAST and DAST-20 is ≥ 6 (Gavin, Ross & Skinner, 1989; Skinner & Goldberg, 1986), ≥ 3 in the DAST-10 (Skinner, 1982), and either 6 or 7 in the DAST-A (Martino, Grilo & Fehon, 2000). The DAST can be administered through paper and pencil or computerized versions (Martino et al., 2000).

Positive Features

- The DAST is brief to administer and is easily scored. A general cut-off score of 6 is used with the DAST. Other versions of the DAST employ cut-off scores varying 3–7 and allow for clinical judgment in determining appropriate cut-offs (Staley & El-Guebaly, 1990; Yudko, Lozhkina, Fouts, 2007)
- The DAST has been found to be more effective than several other drug screening instruments in identifying drug use disorders among offenders (Peters et al., 2000)
- The DAST-10 has good convergent validity with the SCID in detecting alcohol problems and shows incremental validity over the SCID alone (Maisto, Carey et al., 2000; Maisto, Conigliaro et al., 2000)
- The DAST-10 and DAST-20 are related to alcohol, drug, and psychiatric measures, supporting its concurrent validity across different populations and age groups (Yudko et al., 2007; Achenbach, Krukowski, Dumenci, & Ivanova, 2005; Cocco & Carey, 1998; Gavin et al., 1989; Martino et al., 2000)
- The DAST can distinguish between individuals with primary alcohol problems, those with primary drug problems, and those with both sets of problems (Cocco & Carey, 1998; Martino et al., 2000; Staley & El-Guebaly, 1990; Yudko et al., 2007)
- The DAST-10, DAST-20, and DAST-A can discriminate between people with current substance use disorders, people with past substance use disorders, and people who have never had substance use disorders (Cocco & Carey, 1998; Martino et al., 2000; Yudko et al., 2007)
- The DAST, The DAST-10, DAST-20, and DAST-A have high internal consistency (alphas range .74–.95) and good test-retest reliability (r scores range .71–.89). These instruments also have good sensitivity, specificity, and positive predictive value in detecting drug use disorders across different groups (including offenders) that differ by age, gender, and culture (Carey et al., 2003; Cocco & Carey, 1998; El-Bassel et al., 1997; Maisto, Carey et al., 2000; Maisto, Conigliaro et al., 2000; Martino et al., 2000; McCann et al., 2000; Peters et al., 2000; Yudko et al., 2007)
- The DAST has been found to have a single underlying factor, supporting the unidimensionality of the measure (Yudko, Lozkhina, Fouts, 2007; Skinner, 1982; Staley & El-Guebaly, 1990). The DAST-A and DAST-10 have also been found to be unidimensional measures (Carey et al., 2003; Martino et al., 2000)
- The DAST-20 correlates well with the original DAST-28 (Coco & Carey, 1998) and other measures of substance use (MAST, AUDIT, ASI, Children of Alcoholics Screening Test) across different populations and gender and age groups (Cocco & Carey, 1998; El-Bassel et al., 1997; McCann et al., 2000; Saltstone, Halliwell, & Hayslip, 1994; Staley & El-Guebaly, 1990; Yudko et al., 2007), supporting the convergent validity of the measure

- The DAST-A has been found to be a reliable and valid screening device for use with adolescents in psychiatric settings and includes wording tailored for adolescents (Martino et al., 2000). The DAST-A is more likely to underestimate than overestimate substance use problems

Concerns

- The DAST does not examine the quantity or frequency of recent or past substance use and is limited to screening for drug problems
- The validity of the DAST has not been widely examined among individuals with CODs
- There is some evidence that the DAST may consist of five factors, departing from other findings of the unidimensional nature of the instrument (El-Bassel et al., 1997; Yudko et al., 2007). Several studies also indicate that the DAST-20 and DAST-10 have a multidimensional factor structure (Cocco & Carey, 1998; Saltstone et al., 1994; Skinner & Goldberg, 1986; Yudko et al., 2007)
- Research indicates that the DAST-10 may yield a high number of "false negatives" (McCann et al., 2000)
- Studies of the DAST-A have not extensively examined criterion validity (Martino et al., 2000)
- The DAST-28 has several potentially problematic items (items 7 and 20) that are not highly correlated with the overall DAST score (El-Bassel et al., 1997; Skinner, 1982; Staley & El-Guebaly, 1990; Yudko et al., 2007). Similarly, items 4 and 5 of the DAST-20, DAST-10, and item 20 of DAST-A are not highly correlated with the total score (Cocco & Carey, 1998; Martino et al., 2000; Yudko et al., 2007)
- The DAST may result in underreporting or denial of symptoms due to the face validity of test items (El-Bassel et al., 1997; Skinner, 1982; Yudkho et al., 2007). The DAST-A is susceptible to faking good in adolescent populations (Yudko et al., 2007)
- The DAST is a commercial product, although the cost is quite modest

Availability and Cost

The Drug Abuse Screening Test (DAST) instrument can be obtained by contacting The Addiction Research Foundation, Marketing Department, 33 Russell Street, Toronto, Ontario, Canada M5S-2S1 at (416) 595-6000. Additional information regarding the DAST can be obtained at the following site: http://bit.ly/DAST_inst

The DAST can also be downloaded, with information regarding scoring and interpretation of test scores, at the following site: http://www.projectcork.org/clinical_tools/html/DAST.html

Michigan Alcoholism Screening Test (MAST)

The MAST (Selzer, 1971) is a self-administered screening instrument that consists of 25 items related to drinking behavior, symptoms, and consequences of use. The MAST is a public domain instrument that was developed through funding by the NIAAA. The screen uses a yes/no format to inquire about problematic alcohol use and addiction throughout the lifetime (Toland & Moss, 1989). A total score is used to determine alcohol use severity. The MAST is among the most frequently studied substance use screening instruments in clinical settings (Teitelbaum & Mullen, 2000).

The MAST-short version (SMAST; Selzer, Vinokur, & VanRooijen, 1975) is a widely used 13-item screening instrument that examines symptoms of alcohol use disorders. A brief 10-item version, the bMAST is also available to examine lifetime severity of problematic drinking (Pokorny, Miller, & Kaplan, 1972). This version includes items from the original MAST that were highly discriminative for alcohol use disorders. A computer-administered version of the MAST is also available, as is a version for the elderly (MAST-G; SMAST-G; Blow, Gillespie, Barry, Mudd, & Hill, 1998; Morton, Jones & Manganaro,

1996). The recommended cut-off score for identifying problem drinking with the MAST is ≥ 5 (Selzer, 1971), with the SMAST is ≥ 3, (Selzer et al., 1975), with the bMAST is ≥ 6, (Pokorny et al., 1972), with the MAST-G is ≥5 (Morton et al., 1996), and with the SMAST-G is ≥ 3 (Blow et al., 1998).

Positive Features

- The MAST is available in the public domain, is brief to administer, and requires no training
- The MAST has good sensitivity in justice settings and effectively identifies most incarcerated individuals who have severe alcohol use disorders (Peters et al., 2000). The test-retest reliability of the MAST among offenders is .86–.88 (Conley, 2001; Peters et al., 2000)
- MAST scores are associated with risk for recidivism among male and female DWI offenders (Lapham, Skipper, Hunt, & Chang, 2000)
- The MAST demonstrates good validity and sensitivity to detecting alcohol use disorders among people in psychiatric settings (Teitelbaum & Mullen, 2000). For example, the MAST has good sensitivity (88 percent) and moderately good specificity (69 percent) in identifying severe alcohol use disorders among individuals who have schizophrenia (Searles, Alterman, & Purtill, 1990; Toland & Moss, 1989). The MAST is more accurate in identifying alcohol problems among males with schizophrenia than with females (McHugo et al., 1993). The 1-week test-retest reliability of the MAST in a psychiatric sample is .98 (Teitelbaum & Carey, 2000)
- The MAST has been found to be reliable, to effectively discriminate between problem and non-problem drinkers (Mischke & Venneri, 1987), and to identify alcohol use disorders and excessive drinking problems (Bernadt, Mumford, & Murray, 1984)
- Among elderly male outpatients, the MAST demonstrates good sensitivity (91 percent), specificity (84 percent), adequate positive predictive value (70 percent), and good negative predictive value (96 percent; Hirata, Almeida, Funari, & Klein, 2002)
- The MAST has an average test-retest reliability of .81 across groups that differ by age, gender, race/ethnicity; across different versions of the instrument; and across study samples (Shields, Howell, Potter, & Weiss 2007)
- Conley (2001) found the MAST to be a more valid indicator of addiction than the AUDIT
- The MAST and SMAST have equivalent internal consistency across age, gender, race/ethnicity; different study populations; and translated versions of the instrument (Shields et al., 2007)
- The SMAST-G has good sensitivity (85 percent) and specificity (97 percent; Moore, Seeman et al., 2002)
- Using DSM-III criteria, the SMAST was found to have higher sensitivity than the CAGE or of clinician reports (Breakey, Calabrese, Rosenblatt, & Crum, 1998)
- Accuracy for the SMAST tends to improve when individuals are queried about alcohol use problems within the past year rather than over the lifetime (Zung, 1984)
- The SMAST-G has moderate sensitivity (71 percent) and good specificity (81 percent) among the elderly (Moore, Seeman et al., 2002), and an optimal cut-off score of 6 has been identified for use with this population (Beullens & Aertgeerts, 2004)
- The bMAST has been validated in two treatment-seeking samples of alcohol users and contains two factors (perception of drinking and consequences of drinking). The bMAST is moderately correlated with the AUDIT and is as effective as the AUDIT in identifying alcohol use severity (Connor, Grier, Feeney & Young, 2007)
- The bMAST has high specificity and positive predictive value among people

who have alcohol use disorders (Soderstrom et al., 1997) and in hospital samples (Hearne, Connolly & Sheehan, 2002)

Concerns

- The MAST is limited to screening for alcohol problems and does not examine the quantity or frequency of alcohol use
- The MAST lacks a time frame for responses. As a result, positive scores do not necessarily indicate a current alcohol problem
- The MAST was not one of the most effective screening instruments in identifying severe substance use disorders among prisoners (Peters et al., 2000)
- Both the MAST and SMAST tend to have greater sensitivity than specificity and thus misidentify individuals as having substance use disorders (Conley, 2001)
- The MAST has only moderate specificity in psychiatric settings (Teitelbaum & Mullen, 2000) and has low specificity in justice settings (Peters et al., 2000)
- Weights for MAST items were not empirically derived, and items related to drug arrests and liver problems detract from the unidimensionality of the measure (Thurber, Snow, Lewis & Hodgson, 2001)
- Among DUI offenders, MAST scores are only moderately correlated with substance use disorders (Conley, 2001)
- The MAST is not as effective in detecting alcohol problems among men (Teitelbaum & Mullen, 2000)
- In psychiatric and treatment settings, the SMAST underestimates alcohol problems among women (Breakey et al., 1998)
- The SMAST is less sensitive in community treatment samples relative to primary care samples (Chan, Pristach, & Welte, 1994). The bMAST also has low sensitivity in a hospital admissions sample (Hearne et al., 2002)
- Use of the MAST may be problematic for people who have schizophrenia and who have a tendency to answer positively when asked about hallucinations associated with heavy drinking, even when such phenomena are unrelated to alcohol consumption (Toland & Moss, 1989)
- The MAST has wide variability in internal consistency (.43–.93). Fourteen studies report internal consistencies of less than .80, and there is significant heterogeneity in these estimates (Shields et al., 2007). The MAST may produce higher internal consistency estimates in males than females (Shields et al., 2007). Internal consistency of the MAST may be higher among clinical versus nonclinical samples (Shields et al., 2007)
- The bMAST may not be effective in assessing current alcohol consumption, withdrawal symptoms or tremors (Connor et al., 2007)

Availability and Cost

The MAST can be downloaded at no cost at the following site, which includes additional information about the tool: http://bit.ly/MAST_inst

Screening, Brief Intervention, Referral to Treatment–SBIRT

The Screening, Brief Intervention, and Referral to Treatment (SBIRT) process is not an individual screening tool but involves an integrative approach towards screening, intervention, and referral to treatment services that was designed for use in primary health care settings and funded by SAMHSA. The SBIRT approach recommends use of an evidence-based substance use screening instrument, and SAMHSA grantees that have implemented this approach have been required to use the ASSIST screening instrument. However, in general, the SBIRT approach does not specify a particular substance use screening instrument, and a number of instruments reviewed in this section could be potentially used for this purpose. Although designed for use in health care settings, the SBIRT approach can be readily adapted for use in justice settings in which there is a high volume

of offenders screened who are in potential need of treatment services. The SBIRT approach has been widely implemented across the United States and is now a reimbursable service through Medicaid and Medicare in many states.

The SBIRT approach was intended to reduce risk for substance use disorders through early identification, early intervention, and triage to treatment. The approach involves a brief (5–10 minutes) universal screening for indicators of substance use disorders; a seamless transition between screening, brief interventions, and brief substance use treatment; and triage to more intensive and specialized treatment services, if needed. The four steps of SBIRT include (1) screening, (2) brief intervention, (3) brief treatment, and (4) referral to a range of more intensive treatment services (SAMHSA, 2011).

The SBIRT model endorses use of evidence-based substance use screening instruments that can be used across a broad range of populations and settings (e.g., primary care, trauma centers) and that can identify risk levels (e.g., low, moderate, high) related to substance use severity. These risk levels can be used to identify those in need of a brief intervention, brief treatment, and referral to more intensive services. SAMHSA recommends that people identified as being of moderate to high risk for substance use disorders may need brief interventions, brief treatment, and referral for intensive services. Commonly, SBIRT screening tools include the ASSIST, the AUDIT, the CAGE, and the DAST. Prescreening instruments such as the NIDA Quick Screen or the AUDIT-C are often used to identify people who may have significant substance use problems, prior to administration of a more in-depth screening instrument to determine the need for a comprehensive assessment related to substance use disorders.

Positive Features

- SBIRT combines screening for alcohol and other drugs, and those screened as positive are referred for brief intervention or treatment, based on the risk level as determined by substance use severity. The approach uses an integrated model to provide graduated levels of services for people who have varying needs for substance use treatment (Babor et al., 2007)
- SBIRT effectively identifies those who are at risk for substance use problems in primary care settings. People may not be seeking help for substance use problems in these settings, and thus, SBIRT provides a unique set of early intervention and prevention services (SAMHSA, 2011)
- SBIRT provides significant public health savings ($3.81 for every $1 spent; Fleming et al., 2002; Gentilello, Ebel, Wickizer, Salkever & Rivara, 2005)
- SBIRT has been adapted in justice settings, using TICs (Targeted Interventions for Corrections; Joe et al., 2012; Knight, Simpson, & Flynn, 2012), which integrate screening tools such as the TCU scales and the ASI for use in referral to treatment and treatment planning. The TIC system implements a battery of instruments that are tailored for offenders, including measures of substance use, criminal thinking, motivation and treatment readiness, and psychological functioning. Results are then used to place offenders into brief interventions that focus on anger management, HIV/sexual health, motivation, and developing positive social networks. The TIC system also includes referral to more intensive substance use treatment (Joe et al., 2012; Knight et al., 2012)
- Across settings (i.e., primary care, hospitals, public and rural health care offices, inpatient, and outpatient clinics) and use of different universal screening tools (i.e., AUDIT, CAGE, DAST), the SBIRT approach has effectively referred those who screen positive for substance use problems at baseline (17–40 percent) to either a brief intervention (13–70 percent), brief treatment (2–14 percent), or to more intensive treatment (4–16 percent),

resulting in over 63 percent receiving some type of treatment (Madras et al., 2009)

- SBIRT interventions that involve referral to diverse service settings (e.g., trauma centers, emergency rooms, primary care clinics) and that use a range of different screening instruments have yielded significant reductions in substance use over a 6-month follow-up period. These results are consistent across different levels of substance use severity and across age, gender, and race/ethnicity groups (Madras et al., 2009)
- Other studies have shown similarly positive results for screening and brief interventions for individuals who use different types of substances (Bernstein et al., 2005; Copeland, Swift, Roffman & Stephens, 2001; McCambridge and Strang, 2004; Humeniuk et al., 2008; Madras et al., 2009; Schermer, Moyers, Miller, & Bloomfield, 2006; Soderstrom et al., 2007)
- In a study of people screened as having moderate risk for substance use disorders by the ASSIST, people randomly assigned to receive a brief intervention had significantly lower substance use (60 percent reduction) in contrast to a comparison group. These effects did not vary by age or education level (Humeniuk et al., 2008)
- The ASSIST appears to be one of the most comprehensive substance use screens that is used in the SBIRT system, as the instrument addresses different types of substances and different levels of substance use. The ASSIST and subsequent brief interventions are relatively easy to administer (SAMHSA, 2011). Additionally, national and international organizations have recommended using the ASSIST (and the AUDIT), including NIDA, SAMHSA, and WHO
- SBIRT has good potential for identifying people who misuse prescription drugs and in promoting abstinence over a 6-month follow-up period (Office of National Drug Control Policy & SAMHSA, 2012)
- SBIRT is reimbursable through Medicaid, Medicare, and third party insurers in many states (Madras et al., 2009; ONDCP & SAMHSA, 2012)
- SBIRT may also be effective for adolescents who are at risk for substance use disorders (Bernstein et al., 2009; D'Amico, Miles, Stern & Meredity, 2008; Spirito et al., 2004)
- The SBIRT system has produced effective outcomes related to physical and mental health, employment, housing, and IV drug use (ONDCP & SAMHSA, 2012; Madras et al., 2009)
- Use of the SBIRT approach has led to a reduced number of arrests within a 30-day period (ONDCP & SAMHSA, 2012)

Concerns

- SBIRT services have been studied most extensively in primary care and hospital settings, and have not been as carefully examined within justice populations
- Those who receive brief interventions for opioid use disorders based on the ASSIST screening do not always experience significant reductions in substance use or have lower scores on substance use screening instruments over time (Humeniuk et al., 2008). Other studies have not detected changes in substance use among those receiving the SBIRT brief interventions (Marsden et al., 2006). Some reductions in substance use have been identified among comparison groups who received no intervention
- SBIRT may provide different outcomes for those with alcohol problems, as studies have found inconsistencies in response rates, severity of use, and intervention outcomes (Babor, Steinberg, Anton & Del Boca, 2000; Madras et al., 2009; Saitz et al., 2007). For example, Saitz and others (2007) report that people with severe alcohol use disorders who received brief SBIRT interventions did not show a significant reduction in alcohol use relative to a comparison group

- Substance use screening generally employs self-report screening instruments, which may not be as accurate as clinical interviews or the use of self-report instruments in combination with drug testing (Vitale, van de Mheen, van de Wiel, & Garretsen, 2006)
- Additional research is needed to examine the stability of SBIRT-related reductions in substance use over time during follow-up periods of greater than 6 months (Madras et al., 2009)
- SBIRT studies with adolescents have yielded inconsistent results in reducing substance use and are compromised by several methodological problems (Bernstein et al., 2010; Spirito et al., 2011)

SBIRT Resources

Several resources for developing and implementing an SBIRT approach for screening, brief interventions, and referral to treatment are provided at the following sites:

http://www.samhsa.gov/sbirt

http://www.drugabuse.gov/publications/resource-guide

http://www.samhsa.gov/sbirt

http://www.cdc.gov/ncbddd/fasd/documents/alcoholsbiimplementationguide.pdf

Billing codes for SBIRT service are available at the following sites:

http://www.wiphl.org/uploads/media/SBIRT_Manual.pdf

Simple Screening Instrument for Substance Abuse (SSI)

The Simple Screening Instrument for Substance Abuse (SSI; CSAT, 1994) is a 16-item screening instrument that examines symptoms of severe alcohol and drug use disorders that have been experienced during the past 6 months. The instrument was developed by SAMHSA's Center for Substance Abuse Treatment (CSAT) through selection of items from eight existing screening instruments and from the DSM-III-R. The SSI examines five domains related to severe substance use disorders: (1) alcohol and drug consumption, (2) preoccupation and loss of control, (3) adverse consequences, (4) problem recognition, and (5) tolerance and withdrawal. The SSI can be self-administered or provided through an interview. The recommended cut-off score for identifying alcohol or other drug (AOD) disorders is ≥ 4 (CSAT, 1994).

Positive Features

- The SSI is brief to administer and can be easily administered and scored by nonclinicians, without the need for training
- The SSI is available at no cost
- The SSI is one of the most frequently used substance use screening instruments within state correctional systems (Moore & Mears, 2003) and is widely used in other justice settings (DeMatteo, 2010; Knight, Simpson, & Hiller, 2002; Moore & Mears, 2003; Peters et al., 2004; Taxman, Young et al., 2007)
- In a study comparing the psychometric properties of several screening instruments in correctional settings, the SSI was found to be one of the most effective instruments in identifying severe substance use disorders (Peters et al., 2000)
- The SSI had the highest sensitivity (87 percent) and overall accuracy (84 percent) of the several substance use screening instruments examined in a prison-based study and also has good specificity (80 percent; Peters et al., 2000)
- The SSI functions as intended as a unidimensional measure (Boothroyd, Peters, Armstrong, Rynearson-Moody & Caudy, 2013)
- The SSI has good convergent validity with other substance use measures among justice-involved individuals (O'Keefe, Klebe & Timken, 1999)

- The SSI has good convergent validity, and at a cut-off score of 4, has moderate to large effect sizes in identifying people who need substance use treatment, those who have used substances in the past month, those reporting functional deficits, and those who have lower levels of "quality of life" (Boothroyd et al., 2013)
- The SSI exhibits good sensitivity (82 percent), specificity (90 percent), positive predictive value (99 percent), and negative predictive value (37 percent) in a Medicaid population. These psychometric properties are not influenced by ethnicity or gender (Boothroyd et al., 2013)
- The SSI has good sensitivity at a cut-off score of 1 in detecting substance use disorders among college students (Kills Small, Simons & Stricherz, 2007) and was correlated with several other validated measures of substance use disorders (i.e., the AUDIT, Rutgers Alcohol Problem Index-RAPI, and Daily Drinking Questionnaire-DDQ)
- The test-retest reliability of the SSI among justice-involved individuals is quite good (.83–.97; O'Keefe et al., 1999; Peters et al., 2000)
- The internal consistency of the SSI is quite good among adolescents (alpha = .83; Knight, Goodman, Pulerwitz, & DuRant, 2000), adult offenders (alpha = .91; O'Keefe et al., 1999), and Medicaid enrollees (alpha = .85; Boothroyd et al., 2013). Good internal consistency is provided across race/ethnicity and gender groups (alphas = 82–.86; Boothroyd et al., 2013)

Concerns

- The validity of the SSI has not been examined among individuals with CODs
- The SSI may not be as effective in identifying alcohol use disorders as the AUDIT (Kills Small et al., 2007)
- The SSI does not examine the quantity or frequency of recent and past substance use

Availability and Cost

The SSI is available free of charge and is described in the following monograph: The Center for Substance Abuse Treatment. (1994). Simple screening instruments for outreach for alcohol and other drug abuse and infectious diseases. *Treatment Improvement Protocol (TIP), Series 11*. Rockville, MD: U.S. Department of Health and Human Services. This publication may be downloaded at http://store.samhsa.gov. Or, call SAMHSA at 1-877-SAMHSA-7 (1-877-726-4727) (English and Español).

The self-report instrument and scoring instructions are available free of charge at the following site: http://www.ncbi.nlm.nih.gov/books/NBK64629/

Substance Abuse Subtle Screening Inventory (SASSI-3)

The SASSI-3 (Miller, 1985) examines symptoms and other indicators of alcohol and drug use disorders and was designed to identify individuals who may need further assessment and diagnosis of these disorders (Lazowski, Miller, Boye, & Miller, 1998). The SASSI-3 includes an initial section consisting of 67 true/false items and 8 subscales that are described as "subtle" indicators of substance use disorders. Although described as "subtle," many of the items refer directly to substance use. A second section of 12 items examines alcohol use, and a third section examines other drug use for a total of 93 items. Five of the subscales from the first ("subtle") section of the instrument and the two subscales derived from the remaining ("face valid") sections are used in determining a yes/no decision regarding the probability of a substance use disorder. The decision rules in making this determination are somewhat different for males and females.

The instrument may be administered via paper and pencil or by computer (Swartz, 1998). The SASSI-A has been developed for use with adolescents. The recommended cut-off score as indicated by the SASSI-3 user's guide for identifying severe substance use disorders among

adults is ≥ 17 with males and ≥ 19 with females (Miller, Roberts, Brooks & Lazowski, 1997).

Positive Features

- Researchers at the SASSI Institute report that the SASSI, SASSI-2 and SASSI-3 (Miller & Lazowski, 1999) have high sensitivity, specificity, and positive predictive value (Lazowski et al., 1998) across a range of settings
- The SASSI adult manual indicates adequate classification rates of substance use disorders (62 percent; Bauman Merta & Steiner, 1999)
- Several studies examining the SASSI-3 (Arenth, Bogner, Corrigan, & Schmidt, 2001; Ashman, Schwartz, Cantor, Hibbard, & Gordon., 2004) indicate adequate sensitivity (72–85 percent), specificity (63–82 percent), positive predictive value (68–76 percent), and negative predictive value (74–84 percent)
- The SASSI demonstrates adequate agreement with the CAGE and the MAST (Laux, Salyers, & Kotova, 2005; Myerholtz & Rosenberg, 1998)
- The SASSI "direct" scales perform relatively well in classifying substance use disorders (84–89 percent) and perform better than the total SASSI score in this regard (Ashman et al., 2004; Clements, 2002; Gray, 2001; Swartz, 1998)
- The SASSI-A scales have demonstrated good construct validity (Stein et al., 2005), and adequate internal consistency (alphas range .66–.74) is reported with the direct scales (Makini et al., 1996; Nishimura et al., 2001)
- In one study, the SASSI-A accurately classified 76 percent of people who did not admit to alcohol and drug use problems (Rogers, Cashel, Johansen, Sewell, & Gonzalez, 1997)
- Studies indicated good 1- and 2-week test-retest reliability and internal consistency for the SASSI's "face valid" subscales (Clements, 2002; Gray, 2001; Laux, Perera-Diltz, Smirnoff, & Salyers, 2005; Laux, Salyers et al., 2005; Lazowski et al., 1998)

Concerns

- The SASSI is a commercial product and is quite expensive in comparison to other substance use screening instruments
- The SASSI was found to be the least effective of eight screening instruments in identifying severe substance use disorders among incarcerated offenders (Peters et al., 2000). The SASSI had among the lowest overall accuracy (60 percent) of the eight substance use screens examined in the study and had the lowest specificity (52 percent) of the five screening instruments that specifically examined drug use disorders, including the Simple Screening Instrument (SSI) and Texas Christian University Drug Screen (TCUDS) that are described in this monograph
- The SASSI does not address a unitary construct and instead examines several underlying factors, in contrast to the intent of the instrument (Gray, 2001; Rogers et al., 1997; Stein et al., 2005; Sweet & Saules, 2003). The SASSI appears to have low internal consistency, reinforcing the concern that it may be measuring several constructs (Myerholtz & Rosenberg, 1998). Several of the SASSI scales appear to measure emotional problems and not substance use (Stein et al., 2005; Sweet & Saules, 2003). In general, it is unclear what the SASSI indirect scales are measuring (Gray, 2001). Confirmatory factor analysis indicates that the SASSI scales and related scoring keys are inconsistent with the factor structure that was obtained using a large offender population (Gray, 2001)
- The SASSI-3 provides 10 subscales; however, research indicates that a 10-factor structure has a poor fit (Gray, 2001). Similarly the SASSI-A provides a 5-factor structure, yet research indicates several differing factor structures for the instrument, with a relatively low amount of variance (33 percent) accounted for by

any of these structures (Feldstein & Miller, 2007; Rogers et al.,1997; Sweet & Saules, 2003)

- The SASSI produces a high proportion of "false positives" among juvenile offenders (68 percent; Rogers et al., 1997) and adult offenders (51 percent; Swartz, 1998), which may be due in part to identification of lifetime substance use disorders
- The SASSI does not examine the quantity or frequency of recent and past substance use
- Scores on the SASSI appear to be significantly affected by gender, education level, or minority status, and there is considerable inconsistency in these scores across different studies (Coll, Juhnke, Thobro, & Haas, 2003; Bauman et al., 1999; Karacostas & Fisher, 1993; Makini et al., 1996; Risberg, Stevens, & Graybill, 1995; Yuen, Nahulu, Hishinuma, & Miyamoto, 2000)
- Racial/cultural minorities may be more likely to be classified by the SASSI as having substance use disorders than other groups (Bauman et al., 1999; Karacostas & Fisher, 1993; Yuen et al., 2000)
- Results of the SASSI may be distorted by comorbid psychopathology, such as conduct disorder (Bauman et al., 1999), depression (Horrigan, Schroeder, & Schaffer, 2000), and trauma (Savonlahti, Pajulo, Helenius, Korvenranta & Piha, 2004)
- In one of the largest samples examined, the SASSI was found to have a sensitivity of only 33 percent (Svanum & McGrew, 1995). The SASSI failed to classify 41–50 percent of those who self-reported drug use in an intake interview (Horrigan & Piazza, 1999)
- The internal consistency of the SASSI-3 is quite variable, with alphas ranging from very low to very high (.27–95) and highest values associated with the "face validity" and "direct" subscales. Other scales show relatively low validity, with alphas ranging .03–.72
- The 1-month test-retest reliability (r score = .36) and 1-week stability (phi = .63) of the SASSI in determining the presence of a substance use disorder is quite low (Myerholtz & Rosenberg, 1998)
- Direct questions related to substance use symptoms are more effective than subtle or indirect approaches used by the SASSI (Gray, 2001; Myerholtz & Rosenberg, 1998; Svanum & McGrew, 1995). The SASSI-3 "subtle" subscales do not correlate well with criterion variables (Clements, 2002) and provide no improvement in classification over direct questions (Clements, 2002; Myerholtz & Rosenberg, 1997; Swartz, 1998). In one study examining the SASSI-A, the "subtle" subscales identified less than half of individuals who openly admitted substance use (Sweet & Saules, 2003)
- The SASSI "subtle" subscales are susceptible to dissimulation, leading to misclassification (Myerholtz & Rosenberg, 1997). They also demonstrate low test-retest reliability (.25–.45; Gray, 2001; Myerholtz & Rosenberg, 1997) and internal consistency (.08; Clements 2002)
- The SASSI may be susceptible to positive impression management (i.e., attempts to minimize substance use in order to avoid social exclusion or other negative consequences; Myerholtz & Rosenberg, 1997)
- Although the SASSI provides treatment recommendations for interpreting scores, there is no empirical evidence to support these interpretations (Feldstein & Miller, 2007)
- The SASSI-3 and SASSI-A are no more effective than several briefer screening instruments in detecting substance use disorders (e.g., CAGE, DAST, MAST; Clements, 2002; Rogers et al., 1997)
- The SASSI-A Correctional (COR) scale does not appear to be related to measures of

criminal activity and thus may be of limited value in predicting recidivism (Stein et al., 2005)

- No studies report internal consistency for the full SASSI-A (Feldstein & Miller, 2007)

Availability and Cost

The SASSI-3 costs approximately $140 for a set of materials that includes the administration manual, a user's guide, a scoring key, and 25 questionnaires and profile sheets. The SASSI-3 is available for purchase at the following site: https://ecom.mhs.com/(S(fyc3pvmieljp5vnkmkvepf45))/product.aspx?gr=cli&prod=sasi&id=overview

Texas Christian University Drug Dependence Screen V (TCUDS V)

The TCUDS V is a 17-item public domain instrument that was derived from a substance use diagnostic instrument (Brief Background Assessment–Drug-Related Problems section) developed by the Texas Christian University, Institute of Behavioral Research as part of an intake assessment for the Drug Abuse Treatment for AIDS-Risk Reduction (DATAR) project, a NIDA-funded initiative evaluating the effectiveness of new treatment intervention strategies (Simpson & Knight, 1998). The TCUDS V provides a self-report measure of substance use problems within the past 12 months, and is based on the DSM-5 criteria for substance use disorders. The instrument provides a brief screen for frequency of substance use, history of treatment, substance use disorder symptoms, and motivation for treatment. A cut-off score of > 4 on the TCUDS V indicates the presence of a moderate substance use disorder, and a score of > 6 indicates a severe disorder.

Positive Features

- The TCUDS V is brief to administer and can be easily administered and scored by nonclinicians, without significant training
- The TCUDS V has been revised to align with the DSM-5 diagnostic criteria for substance use disorders
- The TCUDS V is available at no cost
- The TCUDS is one of the most frequently used substance use screening instruments within state correctional systems (Moore & Mears, 2003; Peters et al., 2004)
- The TCUDS was found to be one of the most effective screening instruments in identifying inmates with severe substance use disorders in a study comparing the psychometric properties of several different screening instruments (Peters et al., 2000)
- The TCUDS had among the highest sensitivity (85 percent) and overall accuracy (82 percent) among several substance use screening instruments examined in a corrections-based study, and also has good specificity (78 percent; Peters et al., 2000)
- The TCUDS examines major DSM diagnostic symptoms of substance use disorders
- TCUDS scores of greater than 5 among prison inmates are associated with increased risk for recidivism (Baillargeon et al., 2009)
- The TCUDS is significantly correlated with the ASI (Pankow et al., 2012), supporting the convergent validity of the instrument
- Test-retest reliability of the TCUDS among incarcerated individuals is quite good (.89–.95; Knight, Simpson, & Morey, 2002; Peters et al., 2000)
- The TCUDS has good internal consistency in different correctional treatment settings (mean alpha = .87; alphas range .84–.89) and across gender (Simpson, Joe, Knight, Rowan-Szal, & Gray., 2012)
- Concordance between self-report and interview information obtained from an earlier version of the TCUDS (Brief Background Assessment) was quite high (Broome, Knight, Joe, & Simpson, 1996)

Concerns

- The validity of the TCUDS V has not been examined among people who have CODs
- The factor structure of the TCUDS has not been well validated, and the instrument may have a different factor structure across populations and levels of substance use severity (Simpson et al., 2012)
- The TCUDS may not be the most effective singular measure for examining alcohol use disorders (Pankow et al., 2012)
- When administering the TCUDS with incarcerated individuals, it may be useful to concurrently screen for deception, as approximately 7 percent of responses may be invalid due to "faking good," and 8 percent of responses may be invalid due to "faking bad" (Richards & Pai, 2003)

Availability and Cost

The TCUDS V and related information about instrument development, scoring, and interpretation can be obtained from the following site: http://ibr.tcu.edu/forms/tcu-drug-screen/

The following site contains a variety of other useful screening and assessment instruments for use in criminal justice and behavioral health settings: http://ibr.tcu.edu/forms/

Recommendations for Substance Use Screening Instruments

Information regarding substance use screening instruments is based on a review of the literature and research examining and comparing the efficacy of these instruments. Factors considered in recommending specific screening instruments include empirical evidence supporting the reliability and validity of the instrument, relative cost of the instrument, ease of administration, and previous use in the justice system. Although summaries of the instruments include research based on the DSM-IV criteria, recommendations are made considering the degree to which instruments align closely with the new DSM-5 criteria and whether they allow for a seamless transition to the new classification system. Recommendations for screening of substance use disorders also include instruments that can be integrated within an SBIRT approach. Based on these considerations, the following screening instruments are recommended to examine substance use disorders:

1. Either the Texas Christian University Drug Screen V (TCUDS V) or the Simple Screening Instrument (SSI) to identify substance use symptoms and substance use severity. The Alcohol Use Identification Test (AUDIT) may be combined with either the TCUDS V or the SSI if a more detailed screening for alcohol use is needed.

(or)

2. The ASSIST, which screens for a wide range of substances (including alcohol, other drugs, and tobacco) and includes a brief intervention component in addition to recommendations for treatment.

Each of these screening instruments requires approximately 5–10 minutes to administer and score.

Screening Instruments for Mental Disorders

A wide range of mental health screening instruments are reviewed in this section. Without use of a formal screening approach, mental disorders are often undetected in criminal justice settings. As a result, staff are less likely to anticipate suicidal behavior and other mental health problems, and the effectiveness of treatment is reduced. Failure to detect mental disorders among offenders also leads to delay in triage to mental health services, behavioral problems that may be attributed to other causes, early dropout from substance use treatment, rapid cycling through community emergency services, and rearrest and reincarceration (Hiller et al., 2011). A wide range of mental health screens

are available for use in the criminal justice system, including several that are in the public domain and downloadable from the internet. The following section describes mental health screening instruments that are widely used in the justice system, that have been validated for use with offenders, or that show significant promise for use with offenders, including those who have co-occurring disorders (CODs).

Screening Instruments for Depression

Beck Depression Inventory-II (BDI-II)

The BDI-II (Beck, Steer, & Brown, 1996) is a 21-item self-report instrument that examines the intensity of depressive symptoms and suicidality. This instrument is one of the most widely used measures of depression. The BDI-II was developed to correspond to DSM-IV criteria of depression and reviews key symptoms, including agitation, difficulty in concentration, feelings of worthlessness, and loss of energy. Elevated scores on items related to suicidal ideation and hopelessness should be attended to carefully, since these items are the most highly predictive of suicidal behavior. The BDI-6 is a recently developed, shorter version of the instrument (Aalto, Elovainio, Kivimäki, Uutela, & Pirkola, 2012; Beck, Ward, Mendelson, Mock, Erbaugh, 1961). Despite its usefulness in screening for depression and suicide, the BDI-II should not be used in diagnosing depression (as reported for the BDI-I; Sundberg, 1987), which requires a more intensive assessment process. The recommended BDI-II cut-off score for identifying depression is ≥ 16 (Beck et al., 1996; Sprinkle et al., 2002). Computerized versions of the instrument are available, as well as a version in Spanish.

Positive Features

- The BDI-II requires minimal training, and can be administered and scored by a nonclinician
- The BDI-II includes scoring instructions and interpretation of different levels of depressive severity to assist in treatment planning
- The BDI-II is clearly and concisely worded, and the measure can be completed in 5-10 minutes
- Only a fifth grade reading level is required to complete the BDI-II
- The BDI-II has been validated for use with adult offenders (Kroner, Kang, Mills, Harris, & Green., 2011)
- The BDI-II has been successfully used as a screening instrument and outcome measure of depression among prisoners (Harner, Hanlon & Garfinkel, 2010; Johnson & Zlotnick, 2008; Gussak, 2006). The instrument has frequently been used with people with substance use disorders and has been found to be useful in the screening and assessment of depression with this population (Buckley, Parker, & Heggie, 2001)
- The BDI-II is correlated with instruments examining both alcohol and drug use and with severity of substance use problems (Dum, Pickren, Sobell, & Sobell, 2008)
- The BDI-II has been validated with diverse cultural populations and has been translated into several languages (Grothe et al., 2005; Penley, Wiebe, & Nwosu, 2003). The instrument has been found to be unbiased in use among ethnic/racial groups (Sashidharan, Pawlow & Pettibone, 2012). The instrument has excellent content, convergent, and divergent validity across different populations, age groups, and gender groups (Arnau, Meagher, Norris, & Bramson 2001; Dum et al., 2008; Krefetz, Steer, Gulab, & Beck 2002; Steer, Beck, & Garrison, 1986; Storch, Roberti & Roth, 2004). Scores on the BDI-II are significantly correlated with other indices of depression, including the Hamilton Rating Scale for Depression (HAM-D, r score = .71) and the Beck Hopelessness Scale (r score = .68)
- Among females offenders, the BDI-II shows good convergent validity with

another measure of depression, the Beck Hopelessness scale (r score = .55). The instrument is also useful in predicting self-harm (Perry & Gilbody, 2009) and in identifying suicidal ideation (Kroner et al., 2011)

- The BDI-II provides a unidimensional construct of depression across cultures (Nuevo et al., 2009; Shafer, 2006), although it reviews several underlying components of depression (e.g., somatic, affective, and cognitive symptoms; Arnau et al., 2001; Dum et al., 2008; Steer, Ball, Ranieri, & Beck, 1999)
- Among people with substance use problems, the BDI-II exhibits good sensitivity (86–96 percent), specificity (86 percent), and negative predictive value (97 percent) in diagnosing depression (Scott et al., 2011; Seignourel, Green, & Schmitz, 2008). Previous studies examining the BDI also indicate moderately good sensitivity (67 percent) and specificity (69 percent) in diagnosing depression among individuals with alcohol problems (Willenbring, 1986)
- Several studies demonstrate high internal consistency within the BDI-II, including those examining female offenders, alpha=.90 (Kroner et al., 2011) and substance-involved populations (alpha=.95; Dum et al., 2008; Buckley et al., 2001). For the Spanish version of the BDI-II, the average coefficient alpha is .91 (range =.89–.93; Wiebe & Penly, 2005)
- The BDI-II demonstrates good test-retest reliability over 1 week (r score =.74–.96; Beck et al., 1996; Leigh & Anthony-Tolbert, 2001; Sprinkle et al., 2002), a finding replicated with the Spanish version of the instrument (Wiebe & Penly, 2005)
- Use of the BDI-6 in the general population indicates good convergent validity with the BDI-II (r score =.88), and higher scores reflect more severe depression or more recent depression. The BDI-6 exhibits good sensitivity (93–80 percent) and specificity (89–70 percent) in identifying current and past diagnoses of depression (Aalto et al., 2012)
- The BDI-6 has good internal consistency (alpha=.83; Aalto et al., 2012)
- A cut-off score ≥1 or 2 in the BDI-6 is recommended for identifying depression within the past 12 months, and a score of ≥ 4 or 5 is recommended for identifying depression within the past two weeks (Aalto et al., 2012)
- The BDI has higher sensitivity (94 percent) and specificity (59 percent) than the Raskin Depression Scale, the Hamilton Depression Rating Scale (HAM-D), and the Symptom Checklist 90-Revised (SCL-90-R; Rounsaville, Weissman, Rosenberger, Wilber, & Kleber, 1979). The BDI-II is also able to distinguish among varying levels of depressive severity (Steer, Brown, Beck, & Sanderson, 2001)

Concerns

- The BDI is not available in the public domain and is fairly costly to purchase
- Higher BDI cut-off scores may be warranted among males with substance use disorders and male prisoners, as studies suggest that these populations have higher levels of depression than other groups (Beck et al., 1996; Boothby & Durham, 1999; Buckley et al., 2001; Steer, Kumar, Ranieri & Beck, 1998)
- First-time offenders tend to have higher scores on the instrument (Boothby & Durham, 1999)
- Further validation of the BDI-II is needed in criminal justice settings. For example, research is needed to explore the diagnostic accuracy (e.g., sensitivity and specificity) of the BDI-6 among offenders and to identify recommended cut-off scores for depression
- The factor structure of the BDI-II among prisoners is somewhat different than in the general population, suggesting that the instrument may measure other components

of depression that are unique to offenders (Boothby & Durham, 1999)

- The BDI-II may have low specificity with substance-involved populations (Seignourel et al., 2008)
- The instrument should not be used as a sole indicator of depression but rather in conjunction with other instruments (Weiss & Mirin, 1989; Willenbring, 1986). Like other screening instruments, the BDI-II is not a diagnostic tool, and elevated scores do not necessarily reflect a major depressive disorder but rather the presence of depressed mood during the past 2 weeks
- Because the BDI measures subjective feelings of depression, it is difficult to discriminate between normal individuals who are experiencing sadness and those individuals who are clinically depressed (Hesselbrock, Hesselbrock, Tennen, Meyer, & Workman, 1983)
- The BDI-II does not differentiate among varying types of mood disorders (e.g., major depressive disorder and dysthymia; Richter, Werner, Heerlein, Kraus, & Sauer, 1998)
- Women score significantly higher than men on the BDI-II, but these gender differences are not reflected across age and racial/ethnic groups. Despite gender differences being acknowledged by the authors (Steer, Beck, & Brown, 1989), only a single set of interpretive guidelines is provided
- Definitions of depression and the experience of depression may differ across countries (Nuevo et al., 2009)
- An alternate version of the BDI-6 includes items (Beck et al., 1961; Bech, Gormsen, Loldrup, & Lunde, 2009) that are based on core features of the Hamilton Depression Scale (HAM-D), including depressed mood, guilt, work inhibition, difficulty making decisions, indecisiveness, irritability, and fatigue (Bech et al., 2009). However, recommended cut-off scores are not provided for this version of the BDI-6

Availability and Cost

The BDI-II can be purchased from Pearson Clinical Assessment at the following site: http://www.pearsonclinical.com/psychology/products/100000159/beck-depression-inventoryii-bdi-ii.html?Pid=015-8018-370

The cost is $79 for one manual and 25 record forms.

Center for Epidemiological Studies–Depression Scale (CES-D)

The Center for Epidemiological Studies-Depression Scale (CES-D) is a 20-item self-report screen that examines the frequency and duration of symptoms associated with depression. Items review symptoms that have occurred during the past week. A 10-item version of the CES-D is also available (Kohut, Berkman, Evans, & Cornoni-Huntley, 1993) and was developed with an elderly population. The CES-D screen can also be administered as a structured interview. The recommended cut-off score in identifying depression is ≥ 16 for the 20-item version of the CES-D (Radloff, 1977) and ≥ 4 for the 10-item version (Irwin, Artin, & Oxman, 1999).

Positive Features

- The original 20-item CES-D is a public domain instrument
- The CES-D takes approximately 5 minutes to administer and 1–2 minutes to score. The instrument does not require professional clinical training to administer or score
- Cut-off scores are available for use with different clinical and nonclinical populations
- The CES-D has been used in criminal justice settings to screen for depression (Bland et al., 2012; Tatar, Kaasa & Cauffman, 2012; Scheyett et al., 2010). Among people with a history of incarceration, the CES-D is strongly correlated with other validated measures of depression (Bland et al., 2012; Tatar et

al., 2012). The CES-D has good internal consistency when used with offenders (alphas=.71–.94; Bland et al., 2012; Tatar et al., 2012). The short form of the CES-D also demonstrates good internal consistency among offenders (Nyamathi et al., 2011)

- The CES-D has been used with substance-involved populations (Khosla, Juon, Kirk, Astemborski & Mehta., 2011; Perdue, Hagan, Thiede, & Valleroy, 2003) and has been found to be suitably effective in detecting symptoms of depression and in measuring change in these symptoms over time (Boyd & Hauenstein, 1997)
- The CES-D has been used with a variety of clinical and nonclinical populations (Atkins, Marin, Lo, Klann, & Hahlweg, 2010 ; Bakitas et al., 2009; Barnes & Meyer, 2012; Giese-Davis et al., 2011)
- The CES-D has been validated for use with different racial/ethnic groups and has been translated into several foreign languages
- The CES-D short forms show good psychometric properties across clinical and nonclinical populations and across gender, race/ethnicity, and different cultures (Al-Modallal, Abuidhail, Sowan, & Al-Rawashdeh, 2010; Carleton et al., 2013; Cheung & Bagley, 1998; Clark, Mahoney, Clark, & Eriksen, 2002; Cole, Rabin, Smith, & Kaufman, 2004; Kohut et al.,1993; Makambi et al., 2009; Milette, Hudson, Baron, & Thombs, 2010; Opoliner, Blacker, Fitzmaurice, & Becker, 2013; Radloff, 1977; Roberts, 1980; Santor & Coyne, 1997; Zhang et al., 2012). The CES-D is strongly correlated with other measures of depression such as the BDI (Cole et al., 2004; Zhang et al., 2012)
- The CES-D contains four factors (somatic, depressed affect, anhedonia, interpersonal problems) that are consistent across clinical and nonclinical populations, gender, and race/ethnicity (Bush, Novack, Schneider, & Madan, 2004; Makambi, Williams, Taylor, Rosenberg, Adams-Campbell., 2009; Shafer, 2006)
- The CES-D has good psychometric properties for use with adolescent and elderly populations (Dozema et al., 2011; Prescott et al., 1998; Sheehan, Fifield, Reisine, & Tennen, 1995; Wancata, Alexandrowicz, Marquart, Weiss, & Friedrich, 2006), and has sensitivity of 74–84 percent, and specificity of 60–74 percent (Haringsma, Engels, Beekman, & Spinhoven, 2004; Prescott et al., 1998)

Concerns

- Offenders and people with substance use disorders may exhibit elevated scores on the CES-D relative to other populations, which may warrant higher cut-off scores in screening for clinical depression (Bland et al., 2012; Khosla et al., 2011; Perdue et al., 2003; Tatar et al., 2012)
- Further validation in justice settings is needed to examine specificity and sensitivity in detecting depression
- The CES-D may be biased by gender (Stommel et al., 1993), and there may be differences in rates of depression by gender, even after accounting for measurement bias (Van de Velde; Bracke, Levecque, & Meuleman, 2010)
- The CES-D short form may contain two underlying factors of negative affect and lack of positive affect (Zhang et al., 2012)
- The CES-D has shown to have from two to four underlying factors across different populations (Al-Modallal et al., 2010; Carleton et al., 2013; Lee et al., 2008; Makambi et al., 2009; Shafer, 2006; Rivera-Medina, Caraballo, Rodriguez-Cordero, Bernal, & Dávila-Marrero, 2010)

Availability and Cost

The CES-D is available at no cost, and can be obtained at the following address: NIMH, 6001 Executive Blvd. Room 8184, MSC 9663, Bethesda, MD 20892-9663; (301) 443-4513. The instrument can also be downloaded at http://www.emcdda.europa.eu/html.cfm/index3634EN.html

General Screening Instruments for Mental Disorders

Brief Jail Mental Health Screen (BJMHS)

The BJMHS was developed through funding by the National Institute of Justice (NIJ) and was validated using a sample of over 10,000 detainees in four jails. The BJMHS was derived from the Referral Decision Scale (RDS), which was designed to aid correctional staff in identifying individuals who have severe mental disorders (Steadman, Scott, Osher, Agnese, & Robbins, 2005). In developing the screen, the total number of RDS items was reduced, several items were rephrased, and the assessed time span for symptom occurrence was changed from lifetime to the past 6 months. The BJMHS consists of six items that examine the occurrence of mental health symptoms for nine DSM-IV diagnoses, including mood disorders and psychotic disorders. The instrument includes two additional items that review prior hospitalization for mental health problems and current use of psychotropic medication. Individuals who endorse two or more items or who indicate either use of psychotropic medication or a history of prior psychiatric hospitalization are classified as needing additional mental disorder screening. The recommended cut-off score for identifying a mental disorder is ≥ 2 (Steadman et al., 2005).

Positive Features

- The BJMHS is available in the public domain
- The BJMHS requires only 5 minutes to administer and includes scoring procedures, cut-off scores, and interpretation regarding the need for further screening of mental disorders
- Little training is required to administer and score the instrument
- The BJMHS has been tested in forensic populations and is readily adaptable for a range of correctional settings. The instrument has been widely used among jail populations (Steadman et al., 2009) and is recognized as an effective tool in identifying severe mental disorders (Ogloff, Davis, Rivers & Ross, 2007)
- Among jail inmates, the BJMHS is equally effective in identifying lifetime diagnosis for a variety of mental disorders, as determined by results from the Structured Clinical Interview for DSM-IV (SCID-I; Eno Louden, Skeem, & Blevins, 2012)
- The BJMHS exhibits adequate sensitivity (64–81 percent), good specificity (76-84 percent) and an acceptable false negative rate (8–15 percent) across gender groups for mental disorders (Eno Louden et al., 2012; Steadman et al., 2009; Steadman et al., 2005)
- The sensitivity and specificity of the BJMHS are similar to those of the K6 instrument (Eno Louden et al., 2012) and the Jail Screening Assessment Tool (JSAT) in identifying severe mental disorders such as schizophrenia, bipolar disorder, and depressive disorder (Baksheev, Ogloff, & Thomas, 2012)
- The BJMHS has adequate internal consistency (alpha=.63; Eno Louden et al., 2012)

Concerns

- Further validation in criminal justice settings is needed to examine the instrument's specificity and sensitivity
- The BJMHS screens only for severe mental disorders and does not address anxiety or personality disorders (Steadman et al., 2009). The absence of items related to anxiety disorders likely diminishes the instrument's sensitivity (Steadman et al., 2009). For example, the BJMHS performs poorly in identifying anxiety disorders among males (Ford, Trestman, Wiesbrock, & Zhang, 2007). Among offenders, the Jail Screening Assessment Tool (JSAT; Nicholls, Roesch, Olley, Ogloff, & Hemphill,, 2005) demonstrates better sensitivity than the BJMHS for any Axis

I disorder, inclusive of anxiety disorders (Baksheev et al., 2012)

- The BJMHS may be more effective for male rather than female inmates, as the rate of "false-negatives" is significantly higher among female inmates (24–35 percent) than male inmates (8–15 percent; Steadman et al., 2005; Steadman et al., 2009). The BJMHS also provides higher "false positive" rates among women in detecting mood and psychotic disorders (Steadman et al., 2005; Steadman, Robbins, Islam, & Osher, 2007)
- In comparison to the Correctional Mental Health Screen-Male (CMHS-M), the BJMHS provides considerably higher rates of "false positives" for the presence of DSM-IV Axis I or II mental disorders among males (48–59 percent, versus 22–29 percent; Ford et al., 2007)
- The K6 appears to have higher sensitivity than the BJMHS (70 percent versus 46 percent) in detecting the presence of a DSM-IV Axis I mental disorder, as determined by the Composite International Diagnostic Interview Schedule-SF (CIDI-SF; Swartz, 2008)

Availability and Cost

The BJMHS may be obtained at no cost at the following site: http://www.prainc.com/wp-content/uploads/2015/10/bjmhsform.pdf

Brief Symptom Inventory (BSI)

The BSI (Derogatis & Melisaratos, 1983) is a 53-item self-report screen for mental health symptoms. The instrument was adapted from its predecessor, the Symptom Checklist 90–Revised (SCL90-R), and is particularly useful in monitoring treatment outcomes and providing a summary of symptoms at a specific point in time. The BSI includes nine Primary Symptom Dimensions (scales), including Somatization, Obsessive-Compulsive, Interpersonal Sensitivity, Depression, Anxiety, Hostility, Phobias, Paranoid Ideation, and Psychoticism. There are also three Global Indices: Global Severity Index (GSI), measuring overall psychological distress; Positive Symptom Distress Index (PSDI), measuring the intensity of symptoms; and the Positive Symptom Total (PST), measuring the number of self-reported symptoms. A shorter version, the Brief Symptom Inventory-18 (BSI-18) can be completed in approximately 4 minutes. The BSI-18 includes three Symptom Dimensions (Somatization, Depression, and Anxiety) and a Global Severity Index (GSI). A profile report is also provided, which presents raw and normalized T scores for each of the Primary and Global Scales. An interpretive report (not available with the BSI-18) provides a narrative summary of symptoms and scale scores. A progress report is available to monitor an individual's progress over time. The recommended cut-off score to identify psychopathology and psychiatric distress for the BSI is ≥ 63 on the GSI (Derogatis, 1993) and the cut-off score for the BSI-18 is ≥ 57 (Zabora et al., 2001).

Positive Features

- The BSI requires only 8–10 minutes to complete, and a sixth grade reading level. The instrument can be administered via paper and pencil, audiocassette, or computer
- The BSI includes scoring instructions, cut-off scores for each scale and for the GSI, and interpretation of cut-off scores in the context of psychological symptoms and distress
- The BSI has been widely used with different populations in assessing psychiatric symptoms and distress, including offenders (Borduin, Schaeffer & Heiblum, 2009; Houck & Loper, 2002; Kroner et al., 2011), nonclinical populations (Kellett, Beail, Newman, & Frankish, 2003), and clinical populations such as people with substance use disorders (Li, Armstrong, Chaim, Kelly, & Shenfeld, 2007; Meredith, Jaffe, Yanasak, Cherrier, & Saxon, 2007; Schwannauer & Chetwynd,

2007; Booth, Leukefeld, Falck, Wang, & Carlson, 2006)

- The BSI is highly correlated with indicators of psychiatric distress among female offenders (Warren, Hurt, Loper, & Chauhan, 2004)
- Over 400 studies examining the reliability and validity of the BSI indicate that it is a suitable alternative to the SCL-90-R (Zabora et al., 2001). These studies demonstrate good evidence of convergent and construct validity with results of diagnostic interviews (Beail, Mitchell, Vlissides, & Jackson, 2013)
- The dimensions of the BSI are highly correlated with those of the SCL-90-R as are the BSI's Global scores (> .90)
- The BSI-18 contains three factors (somatization, depression, and anxiety) that are identified consistently across different clinical populations and cultures (Dura et al., 2006; Recklitis et al., 2006; Wang, Kelly, Liu, Zhang, & Hao, 2013; Wang et al., 2010)
- Both test-retest and internal consistency reliabilities are very good for the BSI's Primary Symptom Dimensions with offenders and treatment-referred populations (Beail et al., 2013; Kellett et al., 2003)
- The BSI has been translated into several languages

Concerns

- The BSI is not a public domain instrument and is relatively costly
- Separate norms are not provided for criminal justice populations
- The BSI does not distinguish between different types of anxiety disorders and instead measures overall anxiety (Derogatis & Savitz, 2000)
- Several studies involving psychiatric and substance-involved clinical populations, college populations, and Latino populations indicate that the BSI does not reflect the nine-factor structure of the SCL-90-R (Benishek, Hayes, Bieschke, & Stöffelmayr, 1998; Derogatis, & Melisaratos, 1983; Hayes, 1997; Prinz et al., 2013; Ruipérez, Ibáñez, Lorente-Rovira, Moro, & Ortet-Fabregat, 2001) and has varying factor structures among the different populations sampled. These findings suggest that the BSI subscale scores should be interpreted with caution. Exploratory factor analyses of the BSI-18 demonstrate inconsistent results with the original study findings that supported use of subscales related to somatization, depression, and anxiety (Derogatis & Savitz., 2000). Several studies indicate that the BSI may be measuring a single factor related to psychological distress (Asner-Self, Schreiber, Marotta, 2006; Daoud &

  & Kokkinos, 2008; Prelow, Weaver, Swenson, & Bowman, 2005)
- The original nine BSI subscales may not be appropriate for use with juvenile offenders, as a six-factor structure better fits the results obtained with this population. Whitt & Howard (2012) suggest that the different BSI factor structure may be due to greater variation in mental disorders among adolescent psychiatric populations, in comparison with adults

Availability and Cost

The BSI can be purchased by a qualified health care professional from Pearson Assessments at the following site: http://www.pearsonassessments.com/tests/bsi.htm

Costs vary depending on the desired formats and additional materials purchased, such as profile forms, scoring forms, and interpretation forms. The required manual, profile forms (50), and answer sheets (50) cost approximately $132.

Correctional Mental Health Screen (CMHS)

The Correctional Mental Health Screen (CMHS; Ford & Trestman, 2005) is a brief self-report

screening tool for mental disorders in correctional settings. The CMHS was developed using a large correctional inmate sample that included men (N = 1,526) and women (N = 670). An original composite screening measure included 56 items that examined DSM-IV Axis I and II disorders. Separate screening versions were developed for male offenders (CMHS-M; 12 items) and female offenders (CMHS-F; 8 items) and consist of dichotomous (yes/no) items. Six items are identical in both versions, and the remaining two to six items are unique to each version of the CMHS. The shortened item pool in the two CMHS screens was found to significantly predict depression; anxiety; PTSD; and DSM-IV Axis II disorders, excluding antisocial personality disorder. Recommended cut-off scores on the CMHS are ≥ 6 and ≥ 5 for males and females, respectively. Response cards are provided that include columns describing staff comments for each item (e.g., "refused to answer" or "did not know the answer") as well as general comments (e.g., "individual was intoxicated").

Positive Features

- The CMHS is a public domain instrument
- Both versions of the CMHS are brief to administer (3–5 minutes; U.S. Department of Justice, 2007)
- The CMHS provides detailed administration instructions, including scoring and interpretation of scores for service referral. For example, recommendations are provided for "routine referral" if the cut-off score is met or if staff have concerns about the respondent's psychological functioning. "Urgent referral" indicates severe emotional problems such as suicide risk
- The CMHS was developed for use in criminal justice settings (Ford & Trestman, 2005)
- The CMHS-F may be more effective in screening for mental disorders among female inmates than other measures developed for use with offenders (see Steadman et al., 2005; Steadman et al., 2007). For example, at a cut-off score of 5, the CMHS-F exhibited higher accuracy in detecting DSM-IV Axis I or II disorders than the BJMHS (62 percent) and had a lower false negative rate (21 percent versus 35 percent; Steadman et al., 2005)
- The cut-off scores for the CMHS-F and CMHS-M effectively differentiate between offenders who have mental disorders and those who do not (Ford et al., 2007; Ford, Trestman, Wiesbrock, & Zhang, 2009)
- At a cut-off score of 6, the CMHS-M exhibits good sensitivity (80–86 percent) and adequate specificity (61–71 percent) in detecting mental disorders, as demonstrated within large samples of male and female inmates (Ford et al., 2007). The specificity and sensitivity of the CMHS are similar for African American and White inmates. In comparison to other screening measures, the CMHS-F has quite high sensitivity in screening for mental disorders among female African American inmates. Overall, these findings support the generalizability of the CMHS among different ethnic/racial groups (Ford et al., 2007)
- Overall accuracy for the CMHS is 75–80 percent in detecting any mental disorder or personality disorder (except ASPD; Ford et al., 2007; Ford et al., 2009)
- A follow-up study validating the CMHS (Ford et al., 2009) showed an improvement in false negative rates on the CMHS-F (25 percent) in detecting mental disorders as compared with findings from the original validation study and relative to the BJMHS (35 percent; Steadman et al., 2005). False positive rates are lower for the CMHS-F in comparison to the BJMHS (8–16 percent) in detecting mental disorders and personality disorders (Steadman et al. 2005; Steadman et al., 2007)
- A key psychometric indicator, Area Under the Curve (AUC) is high for both the CMHS-M (73 percent) and CMHS-F (80 percent), indicating effective identification of mental disorders (Ford et al., 2009)

- The convergent validity of both the CMHS-F and CMHS-M is supported by strong correlations with indices of mental disorders from correctional records. Both forms of the CMHS also exhibit good discriminant validity and are not significantly correlated with non-mental health indicators (e.g., risk for violence, sex offending, education level; Ford et al., 2007)
- Interrater reliability for the CMHS-M and CMHS-F is quite high (Ford et al., 2007; 2009), with kappas for the CMHS-M ranging .66–1.0 and for the CMHS-F ranging .62–1.0
- Internal consistency for the CMHS-M (r score = .76) and CMHS-F (r score= .82) is also quite good (Ford et al., 2007, 2009)
- Test-retest reliability of the instrument was adequate across several studies (Ford et al., 2007; 2009) for both the CMHS-M (r score = .84) and the CMHS-F (r score = .82)

Concerns

- The CMHS-F exhibits lower sensitivity and specificity for mental disorders among female African American inmates at the cut-off score of 6. As a result, lower cut-off scores are recommended (e.g., ≥ 2 or ≥ 3) that increase sensitivity (75–100 percent), but yield rates of specificity that are relatively lower (29–71 percent) than those obtained for White female inmates. In general, the CMHS-F exhibits lower specificity for mental disorders than the BJMHS and the RDS
- Further validation is needed among offender subpopulations
- The false negative rate for mental disorders on the CMHS-M (18–26 percent) is higher than on the BJMHS (5–15 percent; Ford et al., 2007; Steadman et al., 2005)
- The CMHS-M has lower specificity in detecting anxiety disorders than other mental disorders (42 percent; Ford et al., 2007)

Availability and Cost

The CMHS-F and CMHS-M are available for download at no cost. The instruments and accompanying information regarding interpretation, validation, and scoring can be obtained at the following site: https://www.ncjrs.gov/pdffiles1/nij/216152.pdf

K6 and K10 Scales

The K6 and K10 scales were developed for the U.S. National Health Interview Survey to examine psychological distress (Kessler et al., 2003). The K6 is a 6-item screen that was derived from the 10-item K10, and evidence suggests that the K6 is as sensitive in detecting mental disorders as the K10. The six core domains of the screens are nervousness, hopelessness, restlessness, depression, feeling as though everything takes effort, and feelings of worthlessness. The K10 also addresses functional impairment related to mental disorders and examines whether psychiatric symptoms are attributable to medical problems. Both measures identify severe mental illness (SMI), which is defined as meeting psychiatric diagnosis of one of the DSM-IV mood or anxiety disorders, inclusive of significant distress or impairment (Kessler et al., 2003). The K10 has been found to be somewhat more effective than the K6 in identifying anxiety and mood disorders (Furukawa, Kessler, Slade, & Andrews, 2003). Recommended K6 cut-off scores for identifying SMI is ≥ 6 for offenders and ≥ 13 in the general population (Eno Louden et al., 2012; Kubiak, Beeble, & Bybee 2009; Kessler et al., 2002). The K10 is included in the National Comorbidity Survey Replication (NCS-R) and in the national surveys conducted by the WHO's World Mental Health initiative. The scales are available in both interviewer-administered and self-administered forms.

Positive Features

- The K6 and K10 are available in the public domain

- The K6 and K10 are brief and can be easily administered and scored by nonclinicians. Guidelines for scoring and interpretation of the K6 and K10 are available
- The instruments have been translated into several languages and have been shown to have adequate sensitivity and specificity in correctly identifying mental disorders (Carrà et al., 2011)
- Although the K6 and K10 instruments were validated in a general health setting, studies indicate that the measures are useful in criminal justice settings (Swartz & Lurigio, 2005). Lower cut-off scores are used in offender populations in comparison to the general population
- A number of studies have examined the K6 for use with criminal justice populations, people with substance use disorders, and people who have co-occurring disorders and support the effectiveness of the K6/K10 scales with these populations (Hides et al., 2007; Kubiak et al., 2009; Kubiak, Kim, Fedock, & Bybee, 2013; Rush, Castel, Brands, Toneatto, & Veldhuizen, 2013; Swartz, 2008; Swartz & Lurigio, 2005; Swartz & Lurigio, 2006)
- The scales appear to accurately discriminate between individuals who meet criteria for a diagnosis of a mental disorder and those who do not, across large epidemiological samples inclusive of different cultures and age groups (Anderson et al., 2013; Andrews & Slade, 2001; Baggaley et al., 2007; Furukawa et al., 2003; Kessler et al., 2003; Kessler et al., 2010; Patel et al., 2008; Sakurai, Nishi, Kondo, Yanagida, & Kawakami, 2011)
- The K6 shows adequate sensitivity (76–86 percent) and specificity (65–75 percent) in detecting mental disorders among people with substance use disorders (Rush et al., 2013; Swartz & Lurigio 2006) and has similarly good psychometric properties for use with offenders (sensitivity = 62–76 percent; specificity = 86–90 percent) and across gender groups (Swartz, 2008; Eno Louden et al., 2012). The K6 has better sensitivity and specificity than other screening tools, such as the Addiction Severity Index and the Psychiatric Diagnostic Screening Questionnaire (PDSQ; Rush et al., 2013)
- Studies conducted in several different countries indicate that the K6 provides good results related to Area Under the Curve (AUC; 77–89 percent) in detecting mental disorders (Kessler et al., 2010)
- Psychometric properties of the K6 are both consistent and good across socio-demographic subsamples; cultures; and different populations, including offenders and people with substance use disorders (Andrews & Slade, 2001; Eno Louden et al., 2012; Furukawa et al., 2003; Kessler et al., 2002; Kessler et al., 2003; Kubiak et al., 2009; Patel et al., 2008; Rush et al., 2013; Sakurai et al., 2011; Slade, Johnston, Oakley-Browne, Andrews, & Whiteford, 2009; Swartz & Lurigio, 2006)
- The K10 has been used among juvenile offenders as an index of overall psychological distress (Kenny, Lennings, & Munn, 2008)

Concerns

- The K6 may not be as sensitive in detecting specific mental disorders in comparison to other mental health instruments, such as the CIDI (Composite International Diagnostic Interview) and the PHQ-9 (Patient Health Questionnaire), and is intended to identify the general presence of a serious mental disorder (Kessler et al., 2010)
- The K6 may have lower sensitivity in identifying mental disorders in comparison to the BJMHS when different cut-off scores are used. For example, among substance-involved samples, a cut-off score of 13 on the K6 yields sensitivity of 62 percent, in comparison to 76 percent for the BJMHS. However, when a cut-off of 6 is used, the sensitivity of the K6 improves to 76 percent, which is equivalent to that of the BJMHS. Thus, it is important to calibrate

the cut-off scores according to the specific population examined (Eno Louden et al., 2012; Kubiak et al., 2009; Rush et al., 2013)

- The K6 may exhibit a unidimensional factor structure when used in general community samples, while a two-factor structure has been found (representing anxiety and depression) in a treatment-referred clinical sample (Sunderland, Mahoney, & Andrews, 2012).

Availability and Cost

The K6 and K10 scales include interview-administered, self-administered, and translated versions. Information regarding scoring, cut-off scores, and validation research are available at no cost at the following site: http://www.hcp.med.harvard.edu/ncs/k6_scales.php

The Mental Health Screening Form-III (MHSF-III)

The MHSF-III was designed as an initial mental health screening for use with clients entering substance use treatment programs. The 18-item measure contains yes/no questions examining current and past mental health symptoms. Positive responses indicate the possibility of a current problem and should be followed up by questions regarding the duration, intensity, and co-occurrence of symptoms. The following disorders are addressed in the MHSF-III: schizophrenia, depressive disorders, PTSD, phobias, intermittent explosive disorder, delusional disorder, sexual and gender identity disorders, eating disorders, manic episode, panic disorder, obsessive-compulsive disorder, pathological gambling, learning disorders, and developmental disabilities. A 13-item version of the MHSF-III is described in the literature and has equivalent psychometric properties to the 18-item original version (Ruiz, Peters, Sanchez, & Bates, 2009). The preferred mode of MHSF-III administration is via interview, although the instrument can also be self-administered. The recommended cut-off score for identifying mental disorders is ≥ 3 (Sacks et al., 2007b). A qualified mental health professional should review responses to determine whether a follow-up assessment or diagnostic workup and treatment recommendations are needed.

Positive Features

- The MHSF-III is quite brief to administer, requiring approximately 15 minutes
- The instrument was designed for use with individuals who have co-occurring substance use and mental disorders
- English and Spanish versions of the MHSF-III are available
- The MHSF-III has good convergent validity, including strong correlations with reported trauma, and clinically elevated scale scores on the PAI scales (e.g., anxiety, depression, borderline personality features). The MHSF-III also has good discriminant validity, as indicated by clinical scale scores on the PAI (Ruiz et al., 2009). The 13-item version of the MHSF-III demonstrates similarly good psychometric properties (Ruiz et al., 2009)
- In two studies of prisoners who were enrolled in substance use treatment, the MHSF-III showed adequate sensitivity (81–90 percent) and specificity (48–68 percent), with overall accuracy of 73 percent in detecting a mental disorder (Sacks et al., 2007a; Sacks et al., 2007b). In identifying more severe mental disorders, the MHSF-III provides good specificity (89–93 percent) and adequate sensitivity (35–43 percent), with overall accuracy of 75–76 percent across gender groups
- The MHSF-III has outperformed the Co-occurring Disorders Screening Instrument for Mental Disorders (CODSI-MD) and the Modified Mini Screen-MMS (MINI-M) in overall accuracy and sensitivity in detecting mental disorders (Sacks et al., 2007a). These differences are more pronounced among female inmates (Sacks et al., 2007b)
- The MHSF-III demonstrates good internal consistency among jail inmates (alpha = .89; Ruiz et al., 2009)

- The MHSF-III has excellent content validity and adequate test-retest reliability and construct validity (Carroll & McGinley, 2001)
- Test-retest reliability for the MHSF-III over a 1-week period is acceptable (kappas range 63–77 percent) in identifying people with "any" and "severe" mental disorders (Sacks et al., 2007b)

Concerns

- The cut-off scores provided for the MHSF-III vary based on the purpose of screening and are accompanied by different levels of specificity, sensitivity, and overall accuracy (Sacks et al., 2007a, 2007b)
- The MHSF-III may not be as sensitive as the CODSI-MD in detecting mental disorders among prisoners involved in substance use treatment, because cut-off scores may provide fairly low sensitivity in identifying "any" mental disorder (43–51 percent; Sacks et al., 2007a, 2007b) and "severe" mental disorders (48 percent; Sacks et al., 2007b)
- There is only a moderate amount of published research examining the MHSF-III, and further reliability and validity testing is needed in criminal justice settings. When used with inmates, there are several items within the MHSF-III that detract from internal consistency, and some items may also be difficult to understand among this population (Ruiz et al., 2009)

Availability and Cost

The MHSF-III is available to download at no cost at the following site: http://www.bhevolution.org/public/screening_tools.page

The instrument along with guidelines for administration, interpretation, and scoring is available from the National Center for Biotechnology Information: http://www.ncbi.nlm.nih.gov/books/NBK64187/

Symptom Checklist 90–Revised (SCL-90-R)

The SCL-90-R is an updated version of the Hopkins Symptom Checklist (Derogatis, Lipman, Rickels, Uhlenhuth, & Covi, 1974) and the SCL-90. The instrument provides a 90-item, multidimensional self-report inventory that is designed to assess physical and psychological distress during the previous week. The instrument examines nine major dimensions of psychopathology, including somatization, obsessive compulsiveness, interpersonal sensitivity, depression, anxiety, hostility, phobic anxiety, paranoid ideation, and psychoticism. The Global Severity Index (GSI) for the SCL-90-R provides a summary score of psychopathology. A cut-off score of ≥ 63 on the GSI can be used to identify psychiatric distress and the presence of psychopathology (Derogatis, 1993). The SCL-90-R is available in three formats: paper and pencil, audiocassette, and computerized administration. The BSI is an abbreviated version of the SCL-90-R (53 items), is somewhat easier to score, and includes nine subscales similar to that of the original SCL-90-R. Other short forms of the SCL-90-R (Prinz et al., 2013) include the SCL-27 (27 items, six subscales: depressive, dysthymic, vegetative, agoraphobic, social phobia), the SCL-14 (14 items, three subscales: depression, phobic anxiety, somatization), and the SCL-K-9 (9 items, unidimensional scale reflecting global severity of distress).

Positive Features

- The SCL-90-R and other versions of the instrument require no training and are brief to administer. Interpretative profile reports are available for scoring
- When used to screen for mental disorders in nonpsychiatric populations, and using a cut-off score of ≥ 63, sensitivity and specificity range 73–88 percent and 80–92 percent, respectively (Peveler & Fairburn, 1990)
- In criminal justice settings, the SCL-90-R has been found to outperform other general

measures of psychological functioning among substance-involved populations (Davison & Taylor, 2001; Franken & Hendriks, 2001)

- The SCL-90-R has been frequently used with substance-involved, forensic, and offender populations to assess overall psychiatric distress (Brooner et al., 2013; Chambers et al., 2009; Fridell & Hesse, 2006; Kidorf et al., 2010; Pardini et al., 2013; Sander & Jux, 2006)
- In criminal justice settings, the SCL-90-R and its subscales demonstrate moderate to strong correlations with other validated measures of psychological distress, including the Comprehensive Psychopathological Rating Scale (CPRS; Asberg & Schalling, 1979) and the Present State Examination (PSE; Wing, Cooper, & Sartorius,1974; Wilson, Taylor, & Robertson 1985), supporting the convergent validity of the SCL-90-R
- Among veterans, the 25-item version of the SCL-90-R demonstrates good sensitivity (85 percent) and adequate specificity (65 percent) in identifying people with PTSD (Weathers et al., 1996). Within general medical populations, the SCL-90-R depression scale exhibits good sensitivity (89 percent) and specificity (61 percent; Aben et al., 2002)
- The SCL-90 has good internal consistency, based on results from the normative sample, and alphas that range .77–.90 (Derogatis, Melisaratos, Rickles, & Rock, 1976). Similar results have been obtained with other clinical and nonclinical populations (Olsen, Mortensen, & Bech, 2004; Paap et al., 2011; Schmitz, Kruse, Heckrath, & Tress, 1999)
- The short forms of the instrument (SCL-14, SCL-K-9; SCL-27) are strongly correlated with other measures of psychopathology (BDI) and with the BSI (Prinz et al., 2013), and have favorable psychometric properties (Prinz et al., 2013; Kuhl et al., 2010). For example, the short forms have good internal consistency (alpha > .70), with no differences in internal consistencies across forms and high correlations between subscales (r scores = .85–.98; Prinz et al., 2013)

Concerns

- The SCL-90-R is not a public domain instrument and is fairly costly
- Additional work is needed to establish the validity of the SCL-90-R with subgroups of offenders
- The SCL-90 has poor specificity (39 percent) in diagnosing depression among alcoholics (Rounsaville et al., 1979)
- An examination of the factor structure of the SCL-90-R when used with substance-involved populations suggests a single factor of general psychopathology, indicating that the SCL-90-R fails to differentiate among mental disorders in these settings (Zack, Toneatto, & Streiner, 1998)
- A study involving an outpatient population failed to support the original nine-factor structure proposed by Derogatis et al., 1974, and instead found evidence of a single factor reflecting general psychological distress (Schmitz et al., 2000)
- Other studies indicate that the SCL-90-R is composed of eight rather than nine factors when used in both clinical and nonclinical settings (Arrindell, Barelds, Janssen, Buwalda, & van der Ende, 2006; Arrindell & Ettema, 2003)
- An Item Response Theory (IRT) analysis of the SCL-90-R indicates that 28 items could be removed from the instrument and also suggests a single underlying factor that measures psychological distress (Olsen et al., 2004)

Availability and Cost

The SCL-90-R can be purchased by qualified health care professionals from Pearson Assessments at the following site: http://www.pearsonclinical.com/psychology/

products/100000645/symptom-checklist-90-revised-scl-90-r.html

The required manual, profile forms (50 forms) and answer sheets (50 sheets) cost approximately $132. Costs vary, depending on the desired formats.

Recommendations for Mental Health Screening Instruments

Information regarding screening instruments for mental disorders is based on a critical review of the literature and research comparing the efficacy of these instruments. Factors considered in recommending specific screening instruments include empirical evidence supporting the reliability and validity of the instrument, relative cost of the instrument, ease of administration, and previous use in the justice system. Although summaries of the instruments include research that was based on the DSM-IV criteria, recommendations are made considering the degree to which instruments align closely with the new DSM-5 criteria and that allow for a more seamless transition to the new classification system. Recommended instruments for screening mental disorders are those that address co-occurring mental health issues and are geared specifically towards the criminal justice system. Based on the literature review and these considerations, the following screening instruments are recommended to examine mental disorders:

1. Either the Correctional Mental Health Screen (CMHS-F; CMHS-M)

(or)

2. The Mental Health Screening Form-III (MHSF-III) to address mental health problems

(or)

3. The Brief Jail Mental Health Screen.

Each of these instruments requires approximately 5–10 minutes to administer and score.

Screening Instruments for Co-occurring Mental and Substance Use Disorders

Several screening instruments have been developed that address both mental and substance use disorders. These screening instruments differ in the scope and depth of coverage of co-occurring disorders and in the amount of research support for their validity and use in criminal justice settings. Two of these screens (GAIN-SS, MINI-S) are linked with "families" of screening and assessment instruments, and these larger sets of instruments are described in another section, entitled "Assessment and Diagnostic Instruments for Co-occurring Mental and Substance Use Disorders."

The Behavior and Symptom Identification Scale (BASIS-24)

The BASIS-24 is a 24-item self-report measure used to identify a wide range of mental health symptoms and problems. The instrument examines the degree of difficulty experienced during the previous week across six domains of functioning: depression and functioning, interpersonal relationships, self-harm, emotional lability, psychosis, and substance use. The BASIS-24 was derived from its predecessor, the BASIS-32, to provide a brief, yet comprehensive screen of mental health symptoms and psychosocial functioning that can be used over time to examine changes in mental health status. The BASIS-32 assesses both functional domains (self-understanding, daily living skills, interpersonal relations, role functioning, impulsivity, substance use) and psychopathology (mood disturbance, anxiety, suicidality, and psychosis). Items on both measures are rated on a five-point scale (0 = no difficulty and 4 = extreme difficulty). Both measures include scoring and interpretive reports that indicate the severity of problems (none, a little, moderate, quite a bit, extreme) according to the symptom area. Both versions require a scoring algorithm, and can be scored by hand or by use of computerized

software. The software provides summary scores and domain-specific scores, with higher scores indicating greater symptom severity. Both the BASIS-32 and BASIS-24 application guides provide scoring instructions and interpretation that include cut-off scores that distinguish between clinical and nonclinical samples.

Positive Features

- The BASIS-24 requires 5-15 minutes to complete and can be administered via interview, self-report instrument, or computer
- Only a fifth-grade reading level is required, and the instrument can be administered by paraprofessionals
- The BASIS has been translated into Spanish
- An internet-based scoring tool (Webscore) is available that provides scoring of the BASIS-24 and a summary of results
- Both the English and Spanish versions of BASIS-24 can be used to reliably measure change in symptoms (Eisen, Gerena, Ranganathan, Esch, & Idiculla, 2006; Eisen, Normand, Belanger, Spiro, & Esch, 2004) and have been used with populations that have mental and/or substance use disorders (Goodman, McKay, & DePhilippis, 2013)
- The instrument has been widely used in identifying and monitoring mental health problems and outcomes among populations that have CODs (Deady, 2009; Matevosyan, 2010), including veterans (Fasoli, Glickman, & Eisen, 2010; Slattery, Dugger, Lamb, & Williams, 2013) and those mandated to treatment (Livingston, Rossiter, & Verdun-Jones, 2011)
- The BASIS-32 has also been used with offender populations (Cosden, Ellens, Schnell, Yamini-Diouf, & Wolfe, 2003)
- Several studies provide support for the convergent, divergent, and concurrent validity of the BASIS-32 and the BASIS-24 (Eisen, Dickey, & Sederer, 2000; Eisen et al., 2004). The BASIS-24 has better validity and reliability compared to the BASIS-32 (Eisen et al., 2006)
- The BASIS-24 has better reliability and validity in detecting substance use disorders than the BASIS-32 (Eisen et al., 2004)
- Convergent validity of the BASIS-24 among inpatients and outpatients and across ethnic/racial groups is supported by high correlations with other measures of mental health (Eisen et al., 2006), such as the Short Form Health Survey (SF-12) and the Global Assessment of Functioning (GAF). The BASIS-24 also yields elevated subscale scores for depressive functioning, psychotic symptoms, alcohol and drug use, and emotional lability among people diagnosed with depression, psychosis, substance use disorders, and bipolar disorders (Eisen et al., 2006)
- In a psychiatric sample of people diagnosed with depression, the BASIS-24 subscales of depression functioning, emotional lability, and self-harm are highly correlated with measures of depression (CES-D), worry (Penn State Worry Questionnaire; Meyer, Miller, Metzger, & Borkovec, 1990), emotional lability, and substance misuse, (Kertz, Bigda-Peyton, Rosmarin, & Bjorgvinsson, 2012) supporting the convergent validity of the measure
- Discriminant validity of the BASIS-24 is supported by studies indicating that inpatients with greater overall psychopathology have higher scores than outpatient samples (Cameron et al., 2007; Eisen et al., 2006) The substance abuse scale, and psychosis scale are also able to identify individuals with substance use problems and psychosis among people in residential treatment, community mental health patients, and primary health care patients (Cameron et al., 2007)
- The Spanish version of the BASIS-24 shows good convergent validity, because the summary score is significantly correlated with other self-reported measures of mental health (Eisen et

al., 2010). The BASIS-24 subscales of depressive functioning, psychotic symptoms, and alcohol/drug use also show significant differences between those who are diagnosed with and without these disorders in an inpatient psychiatric sample. The Spanish version of the BASIS-24 also has good discriminant validity for psychotic and self-harm symptoms (Eisen et al., 2010)

- Statistical analysis indicates a good fit for the six BASIS-24 subscales among inpatient and outpatient samples, and across ethnic groups (Eisen et al., 2006, 2010)
- The BASIS-24 and its subscales have good internal consistency across racial/ethnic groups, clinical psychiatric populations, primary care populations , and general populations (alphas > .70; Cameron et al., 2007; Eisen et al., 2006; Kertz et al., 2012; Livingston et al., 2011)

Concerns

- The BASIS instruments have not been extensively examined within criminal justice settings
- The measure was originally designed to assess treatment outcomes and to increase consumer involvement in care, and not necessarily for diagnostic purposes
- The BASIS-32 impulsivity, substance abuse, and psychotic symptoms scales may not be sensitive to change over time (Russo et al., 1997; Trauer & Tobias, 2004)
- The BASIS-24 subscales and summary score may not effectively distinguish between inpatients and outpatients among African American and Latino populations, as no significant differences in scores were found between these treatment populations. The BASIS subscales of emotional lability may not be able to distinguish between those with and without bipolar disorder for these same racial/ethnic groups, across inpatient and outpatient settings (Eisen et al., 2006)
- The Spanish version of the BASIS-24 may have poor discriminant validity for subscales of emotional lability and interpersonal relationships (Eisen et al., 2010)
- The BASIS-24 demonstrates poorer test-retest reliability for inpatient samples, particularly on subscales related to interpersonal relationships, emotional lability, and alcohol/drug use, as indicated by intraclass correlation coefficients (ICCs) of .43–.89 (Eisen et al., 2010)

Availability and Cost

The BASIS-24 instrument is available from McLean Hospital at the following site: http://www.ebasis.org/basis24.php

The cost of the BASIS-24 is based on the number of sites licensed to use the instrument. There is an annual fee of $300 for the first site, $100 for the second site, and $50 for the third site.

Staff at McLean Hospital can also be contacted for information regarding the BASIS-24 at spereda@mcleanpo.mclean.org or (617) 855-2424.

The BASIS-32 instrument can be downloaded free of charge at the following site, but materials do not include interpretation or scoring information: http://infotechsoft.com/products/aspect_forms.aspx?formID=BASIS-32

Centre for Addiction and Mental Health–Concurrent Disorders Screener (CAMH-CDS)

The CAMH-CDS is a computer-administered questionnaire that screens for 11 mental disorders, including substance use disorders. The instrument was developed to provide a brief assessment for co-occurring disorders and is designed to determine whether DSM diagnostic criteria are likely to be met for both current and past disorders. The CAMH-CDS requires 5–20 minutes to administer, depending on the number of disorders reported. The instrument was validated

using three large substance use treatment-seeking samples.

Positive Features

- The CAMH-CDS requires only minimal mental health training to administer
- Test results can be generated by computer, immediately following administration
- The CAMHS-CDS has good sensitivity (86–92 percent) in identifying mental disorders for a variety of populations. For mood disorders, anxiety disorders, and schizophrenia/schizoaffective disorders, the CAMH-CDS exhibits good sensitivity (78–80 percent) and adequate specificity (56–68 percent; Negrete, Collins, Turner, & Skinner, 2004)
- The CAMH-CDS has excellent test-retest reliability for mood disorder and anxiety disorder modules and has moderately good reliability for the schizophrenia module (kappas range .72–.94; Negrete et al., 2004)

Concerns

- The CAMH-CDS has only limited ability to discriminate among different mental disorders
- Although the instrument has a high level of sensitivity in detecting mental disorders, it has significantly lower specificity (40–74 percent) in both double blind and clinical samples. For example, with disorders and symptom presentations such as mania, bipolar disorder–mania, and schizoaffective mania, the CAMH-CDS exhibits relatively low sensitivity (57–62 percent; Negrete et al., 2004). Using the previous DSM multi-axial system, the CAMH-CDS often does not effectively discriminate between mental disorders and personality disorders
- The criterion measure for validating the instrument was an unstructured clinical evaluation conducted by a group of trained psychiatrists who were asked to indicate whether, in their clinical judgment, certain disorders were present within 2 weeks of the administration of the CAMH-CDS
- The CAMH-CDS has not been widely used or tested with criminal justice populations
- Interrater reliability may be lower for schizophrenia/schizophreniform disorders (kappas range 65–69 percent; Negrete et al., 2004), suggesting that the CAMH-CDS may not correctly classify these disorders
- Test-retest reliability was determined after instructing participants that they would be readministered the instrument, thus potentially compromising the results (Negrete et al., 2004)

Availability and Cost

The CAMH-CDS is currently included in TREAT, an electronic roster of assessment and outcome measures developed by CAMH. A license is required to use the measures stored on TREAT, and further costs may be required to use copyrighted instruments. Information regarding the CAMH-CDS and TREAT may be accessed at the following site: http://www.treat.ca/tools.html

Global Appraisal of Individual Needs (GAIN)

The Global Appraisal of Individual Needs (GAIN; Dennis, Titus, White, Unsicker, & Hodgkins, 2006) includes a set of instruments developed to provide screening and assessment of psychosocial issues related to mental and substance use disorders. Among the available GAIN instruments are the GAIN-Short Screener (GAIN-SS), the GAIN-Quick (GAIN-Q), the GAIN-Initial (GAIN-I), the GAIN-Monitoring (90 Day), and the GAIN-Quick Monitoring. The full set of GAIN instruments is reviewed in the section entitled "Assessment and Diagnostic Instruments for Co-occurring Mental Health and Substance Use Disorders." The following section focuses on the GAIN Short Screener (GAIN-SS).

The GAIN-SS includes 20 items and requires approximately 5 minutes to administer. The instrument is suitable for use with both adults and adolescents. Four subscales of the GAIN-SS address internal disorders (IDS), behavioral

disorders (EDS), substance use disorders (SDS), and crime and violence (CVS). There are low (score of zero), moderate (score of 1–2) and high risk levels (score of > 3), which are used for the individual scales and for the total score or total disorders screener (TDS). The recommended cutoff score for the GAIN-SS is ≥ 3 for identifying a mental disorder on the TDS, for both adults and adolescents (Dennis, Scott, Funk, & Foss, 2005). However, those who score ≥1 on any of the individual scales are likely to achieve a positive diagnosis on the full GAIN assessment instrument for that particular scale. All versions of the GAIN can be administered via clinical interview, computer, paper/pencil, or self-report.

Positive Features

- The GAIN-SS is quite brief to administer and is one of the few available screens that addresses both mental health and substance use problems
- Software is available for scoring and interpretation of the GAIN-SS, with comments provided regarding diagnosis and treatment planning. Personal feedback reports (PFR) are also available, as well as software designed for federal grantees, using the Government Performance and Results Act (GPRA) measures
- Computerized versions of the GAIN instrument are available that facilitate administration and interpretation. Validity reports are also provided that identify inaccurate or missing data
- A wide variety of instrument support services are available through the GAIN Coordinating Center
- The GAIN-SS instrument is available in Spanish
- Two different versions of the GAIN-SS are available that address problems occurring in "the past 12 months" or across different time spans (e.g., "past month," "2–12 months ago," "over a year ago," "never")
- Norms for the GAIN instrument have been developed for adults and adolescents and for different levels of care. Additional norms are available by gender, race/ethnicity, co-occurring disorders, and involvement in the juvenile and criminal justice system
- The GAIN-SS has been widely used as a screening tool for mental disorders among offenders (Balyakina et al., 2013; Friedmann, Melnick, Jiang, & Hamilton, 2008; Sacks et al., 2007b; Zlotnick et al., 2008) and substance-involved populations (Friedmann et al., 2008; Lucenko, Mancuso, Felver, Yakup, & Huber, 2010)
- Mental health diagnostic impressions from the GAIN-SS are highly correlated with independent psychiatric diagnoses, across a range of disorders (Dennis et al., 2006)
- Among offenders, the GAIN-SS cut-off score of 2 shows good sensitivity (82 percent) and overall accuracy (73 percent) for any mental disorder. At a cut-off score of 5, the GAIN-SS shows good specificity (96 percent) for severe mental disorders (schizophrenia, major depression, bipolar disorder) across gender (Sacks et al., 2007b), as determined by the Structured Clinical Interview for Axis I DSM-IV disorders–SCID-I for DSM-IV (First, Spitzer, Gibbon, & Williams, 2002)
- The GAIN-SS has good sensitivity (91 percent) and specificity (92 percent) in identifying mental disorders among adults, as indexed by the full GAIN instrument (Dennis et al., 2006). The GAIN-SS also has high specificity (91–99 percent) and sensitivity (92–100 percent) for identifying internalizing disorders, externalizing disorders, substance use disorders, and crime/violence (Dennis et al., 2006). Similar results have been found among adolescents (Dennis et al., 2006)
- The GAIN-SS is highly correlated with the full GAIN-I and its subscales (Dennis et al., 2006)
- Test-retest reliability of the GAIN-SS is good for any mental disorder and for severe mental disorders, as indexed by respective

agreement percentages of 77 percent and 83 percent (Sacks et al., 2007b)

- Among adolescents, the GAIN-SS and its subscales (IDS, EDS, SDS), in addition to the internalizing and externalizing summary score (IEDS), are highly correlated with other measures of mental health, including DSM-IV disorders, Youth Self-Report syndrome scales, and the CRAFFT Substance Abuse Screening Test, for their respective disorders and symptoms (McDonell, Comtois, Voss, Morgan & Ries, 2009)
- The GAIN-SS demonstrates good sensitivity for the following disorders among adolescents: IDS (100 percent), EDS (89 percent), SDS (88 percent), and IEDS (74 percent), resulting in correctly classifying 75 percent, 65 percent, 88 percent, and 78 percent of respective participant groups on these subscales (McDonell et al., 2009)
- The GAIN-SS SDS subscale yields good agreement with another measure of concurrent validity, the CRAFFT (kappa of .76; McDonell et al., 2009). The GAIN-SS also has good internal consistency among adolescents (alpha = .81; McDonell et al., 2009)

Concerns

- The GAIN-SS is a copyrighted instrument, and requires a license agreement and a separate user agreement, which is relatively costly
- The GAIN web version is distinct from the paper instrument and is quite costly but provides administrative, scoring and interpretive reports
- Further validation of psychometric properties, including predictive utility of diagnoses, is needed in adult offender populations
- The GAIN-SS contains only five items related to substance use and does not include an interval measure of alcohol or drug use frequency
- The GAIN-SS IDS subscale appears to show better specificity at a cut-off score of 5 (compared to the traditional cut-off score of 3) for offenders who have severe mental disorders
- The GAIN-SS cut-off scores vary in adult populations 1–3 to provide optimal specificity and sensitivity of subscales (Dennis et al., 2006)
- Although the authors state that the GAIN's sensitivity is favored over specificity, specificity is quite low for the IDS subscale (26 percent) and for the EDS subscale (19 percent), suggesting that the instrument may have a high rate of "false negatives"
- Test-retest reliability for the GAIN-SS for any mental disorder and for severe mental disorders is relatively low at a cut-off score of 2 (kappas range .38–.49), in comparison to screens such as the Mental Health Screening Form-III and the MINI Neuropsychiatric Interview–Modified, MINI-M (Sacks et al., 2007b)
- Agreement between GAIN-SS IDS and EDS subscales and other validity measures (Youth Self-Report [YSR] internalizing scale, YSR externalizing scale, YSR total problems) is relatively poor, with kappas ranging .08–.46. This indicates that the GAIN-SS may not be examining the same constructs as these other measures
- The GAIN-SS subscales demonstrate poorer internal consistency among adolescents than adults, with alphas ranging .55–.89 (McDonell et al., 2009)

Availability and Cost

The GAIN instrument license can be purchased by emailing the GAIN developer at gaininfo@chestnut.org or by calling (309) 451-7762.

The GAIN instrument can be downloaded in both English and Spanish at the following website, but they are copyrighted: https://chestnut.box.com/v/GAIN-SS-Materials. Information regarding administration, scoring, and interpretation of the GAIN-SS, along with the instruction manual, can

be downloaded free of charge. This website also provides psychometric information across age groups, including scales and variable descriptions for all versions of the GAIN.

Training is available for administration, scoring, and interpretation of the GAIN-SS. Unlimited training is provided for users at a cost of either $150 for 3 months or $500 for 12 months of access. Costs for utilizing the GAIN depend on the number of users within an agency accessing the cloud-based system, a one-time set up fee, and the annual user fee for each authorized user. A quote based on project needs can be requested by email at gaininfo@chestnut.org or by calling (309) 451-7900.

The Mini International Neuropsychiatric Interview (MINI)

The Mini International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) is a 120-item structured diagnostic interview that is used to identify DSM and International Classification of Disease (ICD) mental and substance use disorders. The instrument was designed as a brief diagnostic screening and has been examined in numerous research and clinical settings. The MINI is composed of a family of instruments that includes the MINI, MINI-Screen, the Modified Mini Screen-MMS (or MINI-M), the MINI-Kid, and MINI-Plus. The full set of MINI instruments is reviewed in the section entitled "Assessment and Diagnostic Instruments for Co-occurring Mental Health and Substance Use Disorders." The following section focuses on the MINI-Screen and the MINI-M instruments.

The MINI-Screen refers the examiner to complete a follow-up module for a particular disorder, if the respondent endorses a threshold screening question. If the respondent does not endorse the item, the interviewer moves to the next section. The MINI screen contains 24 items, including items that assess mood disorders, anxiety disorders, drug/alcohol disorders, and psychotic disorders, based on DSM-IV criteria. However, the Modified Mini Screen (MMS) is a 22-item measure that assesses mood, anxiety, and psychotic disorders only. Therefore, the difference between the MINI Screen and the MMS is that the MMS does not include items aimed at screening for drug/alcohol use disorders. Recommended cut-off scores range 6–9 and are interpreted by a clinician (Alexander, Haugland, Lin, Bertollo, & McCorry, 2008).

Positive Features

- Only brief training is required to use the instrument
- In a combined sample consisting of those in alcohol and drug treatment, in primary health care settings, and in community mental health treatment, the Modified Mini Screen (MMS) demonstrates adequate sensitivity (63–82 percent) and specificity (61-83 percent) at cut-off scores of 6–9 for the Structured Clinical Interview for DSM-IV Axis I (SCID-I) diagnoses of mood, anxiety, and psychotic disorders, and 37–57 percent of participants were referred for further assessment. Similar results have been obtained for different gender and race/ethnicity groups (Alexander et al., 2008). In a study involving participants in family assistance programs, the MMS exhibited adequate specificity (63–86 percent) and sensitivity (61–96 percent) at cut-off scores of 6–12, with overall accuracy ranging 76–77 percent for SCID-I diagnoses and 43–58 percent for referral to treatment (Alexander, Layman, & Haugland, 2013)
- The MMS was found to have higher sensitivity and specificity than other screens, such as the Brief Jail Mental Health Screen (BJMHS) and the K-6 (improved sensitivity only over the K-6; Alexander et al., 2008)
- Among offenders, the MINI-M or MMS demonstrates good sensitivity (71 percent) at a cut-off score of 5, with overall accuracy of 69 percent for any mental disorder as indexed by the SCID-I (Sacks et al., 2007b). Findings are similar across gender groups. For severe mental

disorders (schizophrenia, major depression, and bipolar disorder) identified by the SCID-I, at a cut-off score of 10, the MMS/MINI-M exhibits adequate specificity (84 percent) and overall accuracy (70 percent; Sacks et al., 2007b). The MMS has good internal consistency (alphas = .90–.92), and interrater reliability is quite good (92 percent). Test-retest reliability over a period of 1 week was found to be quite high (Alexander et al., 2008, 2013)

Concerns

- Further validation of the MINI-M is needed in offender populations for screening mental disorders
- In comparison to clinical interviews, use of the MINI results in more frequent diagnosis of co-occurring disorders (Black, Arndt, Hale, & Rogerson, 2004)
- The MINI-Screen includes only one question related to alcohol use and one question examining drug use. The instrument does not include an interval measure of frequency or quantity of substance use
- The MINI-M/MMS appears to exhibit poor specificity for any mental disorder (61 percent) at a cut-off score of 5, as determined by the SCID-I, and has poor sensitivity (42 percent) in detecting severe mental disorders at a cut-off score of 10 (Sacks et al., 2007b)

Availability and Cost

The MINI-Screen can be obtained from the developers' website as part of the entire MINI package, inclusive of the MINI-Screen. For $2, the screen may be downloaded up to 2 times; however, a download does not indicate a licensing agreement. If an organization purchases the MINI package inclusive of the MINI-Screen, price varies based on number of uses. For instance, at the time of this writing, 25 administrations is $125.

The MINI package that includes the MINI-Screen can be obtained at the following site: http://harmresearch.org/index.php/mini-international-neuropsychiatric-interview-mini/

Psychiatric Diagnostic Screening Questionnaire (PDSQ)

The Psychiatric Diagnostic Screening Questionnaire (PDSQ) is a 126-item self-administered instrument that can be used for screening and diagnosis of mental disorders (e.g., mood disorders, anxiety disorders, psychotic disorders) and substance use disorders. The PDSQ provides separate subscales for alcohol use disorders and drug use disorders. The PDSQ examines 13 frequently occurring mental disorders and was designed to evaluate recent psychopathology and to provide background information prior to a more extensive diagnostic evaluation. The PDSQ is described in more detail in the section entitled "Assessment and Diagnostic Instruments for Co-occurring Mental and Substance Use Disorders."

Positive Features

- The PDSQ is 126-item measure that addresses 13 of the DSM-IV Axis I disorders and includes a 6-item screen for psychosis
- The PDSQ requires approximately 15-20 minutes to administer
- The PDSQ includes cut-off scores for individual DSM diagnoses, yielding a sensitivity of > 90 percent (Zimmerman & Mattia, 2001b)
- The PDSQ reflects a single underlying dimension, indicating that the instrument examines a unitary construct, with 15 symptom domains that are independent but all contribute to the unitary construct (Gibbons, Rush, & Immekus, 2009)
- With the exception of the psychosis and somatization subscales, the internal consistency of the PDSQ subscales are > .70, with a mean value of .86, (Zimmerman & Mattia, 1999b, 2001a, 2001b; Gibbons et al., 2009)

- Test-retest reliability of the instrument ranges .61–.83, using relatively stringent criteria, with 9 of 15 subscales demonstrating reliability of > .80 (mean of .83) (Zimmerman & Mattia, 1999b, 2001a, 2001b)
- Diagnostic accuracy of the PDSQ is quite good, with sensitivities ranging .80–.90 and specificity .66–78 (Zimmerman & Mattia, 2001b)
- A receiver operating characteristic (ROC) curves analysis demonstrates that the PDSQ predicts diagnoses significantly better than chance, in reference to the SCID-IV (Sheeran & Zimmerman, 2004)

Concerns

- The PDSQ requires significantly more time to administer than other screens for mental disorders
- The PDSQ generates multiple cut-off scores for different mental disorders, and may require more time to interpret than screening instruments that provide uniform cut-off scores for mental disorders
- Results from studies investigating the PDSQ may not be generalizable to other clinical populations, specifically those that include people who have psychosis and other serious mental disorders. Validation studies have been limited primarily to outpatient populations, and further research is needed to examine the psychometric properties of the PDSQ with a broader range of clinical populations
- The PDSQ is not frequently used in the criminal justice system, and there is little validation research involving offenders
- There is poor internal consistency for two of the PDSQ subscales (psychosis, somatization), with alphas < .70. (Zimmerman & Mattia, 2001a, 2001b)
- Positive predictive values for some PDSQ subscales are quite low .30–.32 (Zimmerman & Mattia, 2001b)
- A factor analysis indicated that only 13 of 15 subscales emerged as factors related to the PDSQ, and only 10 of these were aligned with DSM-IV diagnoses. No major factor was extracted for psychosis, and there was little differentiation between panic and agoraphobia disorders, and between somatization and hypochondriasis disorders

Availability and Cost

The PDSQ can be purchased at the following site: http://www.wpspublish.com/store/p/2901/psychiatric-diagnostic-screening-questionnaire-pdsq

The cost to purchase the PDSQ is $136.50 for 25 test booklets, 25 summary sheets, an instruction manual, and a CD containing 13 follow-up interview guides (one for each of 13 disorders).

Recommendations for CODs Screening Instruments

Information describing screening instruments that address both mental and substance use disorders (CODs) is based on a critical evaluation of available instruments and a review of research comparing the efficacy of these screeners. Key factors used in comparing the instruments include empirical evidence supporting both the reliability and validity of the instrument, relative cost of the instrument, ease of administration within the criminal justice settings, and previous use and evidence of effectiveness within the criminal justice system. Although validity indices for screens described in this section are typically based on previous versions of the DSM (e.g., DSM-IV), recommendations regarding instruments are predicated on their alignment with the recently developed DSM-5, allowing for a more seamless transition from DSM-IV to DSM-5. The following is a recommended screening instrument that addresses both mental and substance use disorders:

- The MINI-Screen addresses a range of co-occurring mental and substance use problems. The MINI-Screen requires approximately 15 minutes to administer and score

In addition, separate screening instruments for mental and substance use disorders can be used in combination. The Brief Jail Mental Health Screen (BJMHS) or the Correctional Mental Health Screen (CMHS-F/CMHS-M) can be combined with the Texas Christian University Drug Screen V (TCUDS V). Refer to the sections "Screening Instruments for Mental Disorders" and "Screening Instrument for Substance Use Disorders" for descriptions and availability information.

Screening and Assessment Instruments for Suicide Risk

People with mental disorders account for a majority of completed and attempted suicides (Cavanagh, Carson, Sharpe, & Lawrie, 2003; Nock et al., 2008), and approximately 63 percent of individuals who complete suicide have a substance use disorder (Duberstein, Conwell & Caine, 1994; Conwell et al., 1996; Schneider, 2009). Although mental disorders account for approximately 10 percent of completed suicides, suicide risk increases to 14–19 percent with the presence of a substance use disorder (Office of Applied Studies, 2006). The risk for suicide is seven times higher among people who have two or more disorders (Nock et al., 2009; Rush, Dennis, Scott, Castel & Funk, 2008).

Suicide is a major concern within the criminal justice system, in which inmates have a 6–7.5 times greater risk than the general population (Jenkins et al., 2005). Males account for 93 percent of completed suicides, and among jail inmates, the risk for suicide is highest within the first month of incarceration. In fact, over half of completed suicides in jail occur within the first 2 weeks of incarceration. Among jail inmates, 80 percent of suicides occur within 2 days of a court hearing (Hayes, 2010). Almost half of inmates who commit suicide have substance use problems (Hayes, 2010). In addition, 20 percent of inmates who complete suicide are under the influence of drugs or alcohol. Mental health problems also contribute to suicides in jail; specifically, 38 percent of inmates who commit suicide have mental disorders, and 20 percent have used psychotropic medications (Hayes, 2010).

Although most jails have written policies and procedures regarding assessment of suicide risk, these are not always effective. For example, 77 percent of jail screenings assess suicide risk at intake, but only 31 percent of correctional officer reporting protocols include risk for suicide, and suicide risk is followed up by correctional staff in only 27 percent of cases in which suicide risk is identified (Hayes, 2010). In cases of completed suicide, 37 percent of inmates were assessed for suicide risk by a clinician, and just under half of completed suicides occurred within 3 days of clinical assessments. Although many correctional facilities provide close observation for those deemed to be at risk for suicide, these observational periods are not continuous and are typically of short duration (e.g., 15 minutes at a time; Hayes, 2010). Given the high rates of suicide in criminal justice settings, implementation of evidence-based instruments for screening and assessment of suicide risk is of critical importance.

In order to provide a comprehensive approach to screening and assessment of suicide risk, it is useful to examine two major components: (1) desire, and (2) capability (see description of these factors in the section entitled “Special Clinical Issues in Screening and Assessment for Co-occurring Disorders in the Justice System”). Therefore, suicide risk instruments should address both of these areas. A number of instruments examine the interaction of these two factors in the context of suicide risk, while other instruments examine a broader range of risk factors related to suicide. The following section describes both interview and self-report instruments that examine risk for suicide. Interview approaches typically address not only desire and capability but other risk and protective factors as well. The self-report instruments, although shorter to administer, do not typically address the full range of risk and protective factors. Further information regarding suicide risk factors within the criminal justice system is provided in the section entitled “Special

Clinical Issues in Screening and Assessment for Co-occurring Disorders in the Justice System." As noted previously, all offenders who screen positively for suicide risk should be immediately referred for a more comprehensive assessment to determine the need for treatment services, close monitoring, and other interventions.

Suicide Risk Screening Instruments

The Adult Suicidal Ideation Questionnaire (ASIQ)

The ASIQ (Reynolds, 1991) is a 25-item self-report measure that was adapted from the 30-item Suicide Ideation Questionnaire (Reynolds, 1987). The ASIQ addresses frequency of suicidal thoughts, plans, and preparation for suicide during the past month. Respondents indicate frequency of thoughts on a 7-point scale (0 = never had this thought, 6 = almost every day). Six critical items are included that are best able to discriminate between those who attempt suicide and non-attempters (Reynolds, 1991). A cut-off score of 14 is recommended in clinical samples, and a score of 31 is recommended in community samples (Osman et al., 1999; Reynolds, 1991).

Positive Features

- The ASIQ has been used with offenders (Horon, McManus, Schmollinger, Barr & Jimenez, 2013)
- The ASIQ is correlated with other indices of suicidal ideation, including the Beck Hopelessness Scale (BHS), the Beck Scale for Suicide Ideation (BSS), and Reasons for Attempting Suicide (RASQ). Scores on the ASIQ are negatively correlated with protective factors as identified by the Suicide Risk Assessment Scale (SRAC), supporting the convergent and discriminant validity of the measure with offenders (Horon et al., 2013)
- The ASIQ is able to discriminate between offenders who have multiple suicide attempts and those who have had a single attempt or no attempts, as evidenced by measures assessing the frequency of suicidal ideation and contemplation and the critical items. The ASIQ more effectively predicts multiple suicide attempts than other suicide risk instruments, such as the BSS and RASQ (Horon et al., 2013)
- In a psychiatric sample, the ASIQ is moderately to strongly correlated with other measures of suicidal ideation, including the BSS, the Suicide Probability Scale (SPS), the BHS, the Beck Depression Inventory (BDI), and the Beck Anxiety Inventory (BAI; Bisconer & Gross, 2007)
- Among psychiatric outpatients, the ASIQ items load highly on a factor related to suicidal ideation, as measured by a composite variable of the ASIQ and the Inventory of Depression and Anxiety Scales (IDAS), supporting the convergent validity of the instrument (Naragon-Gainey & Watson, 2011)
- The ASIQ distinguishes between those at risk for suicide and "controls" in a psychiatric sample (Bisconer & Gross, 2007)
- The ASIQ is able to discriminate between those with and without a history of suicide attempts in a psychiatric sample (Osman et al.,1999)
- The ASIQ predicts suicide attempts during a 3 month follow-up period among psychiatric patients who have previously attempted suicide, supporting the predictive validity of the instrument (Osman et al., 1999)
- The ASIQ's area under the curve (AUC) in identifying multiple suicide attempters is quite good (AUC = .80 total scale; AUC = .69 for critical items; Horon et al., 2013)
- The instrument's specificity is quite good in psychiatric samples (78 percent) when compared with historical records of suicidal ideation and behaviors (Bisconer & Gross, 2007)
- A confirmatory factor analysis yields a single factor, indicating that the ASIQ

measures a unitary construct of suicide ideation (Osman et al., 1999)

- Internal consistency of the entire ASIQ is quite good (alpha = .95–.96; Bisconer & Gross, 2007; Horon et al., 2013; Reynolds, 1991), as well as for the critical items (alpha = .85; Horon et al., 2013) among offender and community samples
- The ASIQ's test-retest reliability over a 1-week interval is quite good (r score = .95; Reynolds, 1991)

Concerns

- The ASIQ has not been widely studied in criminal justice settings
- The ASIQ is not a public domain instrument
- Cut-off scores for the ASIQ may vary between clinical and nonclinical populations
- The sensitivity (51 percent) of the ASIQ is lower than use of historical records in identifying suicidal ideation and behaviors in a psychiatric sample (Bisconer & Gross, 2007)

Availability and Cost

The ASIQ can be purchased from Psychological Assessment Resources, Inc. (PAR), at the following site: http://www4.parinc.com/Products/Product.aspx?ProductID=ASIQ#Items

An introductory kit costs approximately $100, which includes 25 copies of the instrument and an administration manual that provides instructions for administration, scoring, and interpretation.

Beck Scale for Suicide Ideation (BSS)

The BSS (Beck & Steer, 1991) is a 21-item self-report scale that examines thoughts, plans, and intent to commit suicide and includes five screening items. The BSS items inquire about the desire to live, suicidal intent, plans and preparation for suicide, and openness about sharing suicidal thoughts with others. Two additional items examine the frequency and severity of past suicide attempts. If the respondent positively endorses item #4 (desire to make an active suicide attempt) or #5 (duration of suicidal ideation), then items 6–19 are also completed. The instrument requires approximately 5–10 minutes to administer and score. Total scores range 0–38, with 0–2 points assigned to each item, and with higher scores indicating a higher risk for suicide.

Positive Features

- The BSS is brief to administer and score
- The BSS has been used with offenders (Horon et al., 2013; Kroner et al., 2011; Lohner, & Konrad, 2006; Palmer & Connelly, 2005; Senior et al., 2007; Way, Kaufman, Knoll, & Chlebowski, 2013)
- Among offenders who have CODs, the BSS has good convergent validity with other measures of suicide risk, including the ASIQ, RASQ, and the SRAC (Horon et al., 2013)
- The BSS and the BSS screening items are able to discriminate between multiple attempters and non-attempters or single attempters and are able to more effectively predict multiple suicide attempts in comparison to other measures of suicide risk, including the ASIQ and RASQ (Horon et al., 2013)
- Among offenders, the BSS is related to other indices of suicide, including suicidal ideation, suicidal thoughts, and past suicide attempts, as measured by the Depression Hopelessness Suicide Screening form, providing support for its convergent validity (Kroner et al., 2011)
- BSS scores for current suicidal ideation among offenders reporting multiple suicide attempts is significantly higher than for those with only one reported suicide attempt, supporting the validity of the BSS among offenders who have mental health problems (Way et al., 2013)
- The BSS area under the curve (AUC) is quite good (.74) as is the AUC for the BSS screening items (.71), in classifying people

who have multiple prior suicide attempts (Horon et al., 2013)

- Studies involving several international offender populations provide support for the convergent and concurrent validity of the BSS (Lohner & Konrad, 2006; Senior et al., 2007)
- Among veterans, the BSS is able to distinguish between those with and without suicidal ideation. The instrument also detects higher rates of suicidal ideation among veterans who have CODs in comparison to those who have mental disorders only, supporting the validity of the BSS (Bahraini et al., 2013). The BSS demonstrates good internal consistency among offenders (alpha = .85; Horon et al., 2013) and has high levels of internal consistency (alpha = .84), temporal stability, and predictive validity when used to make decisions about hospital admissions (Beck, Brown, & Steer, 1997)
- The BSS has better specificity and positive predictive value in identifying suicide risk than the BHS and the BDI (Cochrane-Brink, Lofchy, & Sakinofsky, 2000)
- A computerized version of the BSS is available. In a study comparing computerized self-report, pen and paper self-report, and clinician report, both self-report versions of the BSI correlated highly (r score > .90) with the clinician reports (Beck, Steer, & Ranieri, 1988)

Concerns

- The BSS is not a public domain instrument
- Additional research is needed to determine the psychometric properties of the BSS with offenders who have CODs. The BSS may not be related to prior suicide attempts in some criminal justice samples (Way et al., 2013)
- Mean scores on the computerized self-reported measure are higher than the clinical ratings, indicating that this measure may yield elevated levels of suicidal ideation (Beck et al., 1988)
- Caution should be taken when interpreting BSS suicide risk severity scores, as offenders may not be willing to report suicidal ideation and may underreport the true severity of suicidal thoughts and desires (Way et al., 2013)
- Analysis of the BSS among clinical samples indicates that it may consist of two to four factors (Beck et al., 1997; Beck, Weissman, Lester, & Trexler, 1976; Witte et al., 2006; Kingsbury, 1993; Spirito, Sterling, Donaldson, & Arrigan, 1996). Several studies indicate a three-factor solution but provide ambiguous results about the nature of the factors (Beck, Kovacks, & Weissman, 1979; Steer, Rissmiller, Ranieri, & Beck, 1993). Thus, caution should be exercised when interpreting BSS scores

Availability and Cost

The BSS is commercially available and can be purchased from the Pearson Assessment website: http://www.pearsonclinical.com/psychology/products/100000157/beck-scale-for-suicide-ideation-bss.html

The administration manual costs approximately $7 and provides scoring and interpretation, while a package including 25 forms of the instrument costs approximately $54.

Interpersonal Needs Questionnaire (INQ)/Acquired Capability for Suicide Scale (ACSS)

The Interpersonal Needs Questionnaire (INQ) and the Acquired Capability for Suicide Scale (ACSS; Van Orden et al., 2012) are two self-report instruments that are administered as a single screening protocol. These are based on the Suicide Risk Decision Tree approach. These instruments provide a direct measure of both suicidal desire and capability. The INQ contains two subscales, one that assesses feelings of burdensomeness (seven items) and another that assesses lack of belonging (five items). The ACSS measures suicide capability (five items). Higher

scores on the ACSS reflect greater suicidal desire and capability and greater suicide risk. Although the INQ and ACSS can be used independently, in combination they provide a comprehensive measure of suicide risk. The INQ/ACSS has not been evaluated in criminal justice settings but shows significant promise in studies of community samples.

Positive Features

- The INQ is a public domain instrument
- The INQ is brief to administer and easy to score
- Among psychiatric outpatients, INQ scores for depression and feelings of burdensomeness and ACSS scores for acquired capability are correlated with clinician-rated risk of suicide, and INQ scores are also associated with suicide capability and desire (Van Orden, Witte, Gordon, Bender, & Joiner, 2008), supporting the convergent validity of the instrument (Van Orden et al., 2008)
- As detected by the INQ, both feelings of burdensomeness and lack of belonging are associated with increased PTSD symptoms and poor mental health in a military sample, supporting the concurrent validity of the instrument (Bryan, 2011)
- Among people involved in substance use treatment, INQ scores related to feelings of burdensomeness and lack of belonging predict risk of suicide attempts, supporting the validity of the instrument (Connor, Britton, Sworts, & Joiner, 2007)
- INQ/ACSS scores for feelings of burdensomeness and suicidal capability are correlated with scores on the Suicidal Behavior Questionnaire-Revised (SBQ-R; Osman et al., 2001). The combination of these two factors is also correlated with suicidality, providing additional support for the convergent validity of the INQ/ACSS (Bryan, Clemens, & Hernandez, 2012)
- The INQ/ACSS is correlated with suicidal ideation among college students, as measured by the Depressive Symptom Inventory–Suicidality Subscale (Davidson, Wingate, Rasmussen, & Slish, 2009)
- Both subscales of the INQ (feelings of burdensomeness, lack of belonging) are correlated with alcohol problems among college students (Lamis & Malone, 2011)
- Higher depression and social anxiety in college students are correlated with feelings of burdensomeness, supporting the construct validity of the INQ among people who have mental disorders (Davidson, Wingate, Grant, Judah, & Mils, 2011)
- The two-factor structure of the INQ (feelings of burdensomeness, lack of belonging) is supported by a study involving a military sample (Bryan, 2011)
- Internal consistency of the INQ and ACSS is quite good, with alphas for the INQ ranging .83–.94 and alphas for the ACSS ranging .83–.85 (Bryan et al., 2012; Nademin et al., 2008)

Concerns

- As noted previously, there has been little research examining the INQ/ACSS with offender populations
- The INQ/ACSS does not yield a threshold or cutoff score indicating high risk for suicide
- For young adults who report suicidal ideation, the interaction of feelings of burdensomeness and lack of belonging does not predict suicide attempts, thus introducing concern about the validity in using the INQ/ACSS with this population (Joiner et al., 2009)
- In a military sample, suicide capability is related to lack of belonging but not feelings of burdensomeness, suicidality scores, or symptoms of depression. Thus, suicide capability should not be used as an independent measure to predict risk of suicide with this population (Bryan, Cukrowicz, West, & Morrow, 2010)

Availability and Cost

The INQ/ACSS is a public domain instrument and is available at the following site: http://psy.fsu.edu/~joinerlab/measures/ACSS-FAD.pdf

Suicide Risk Assessment Instruments

Suicide Risk Decision Tree Interview

The Suicide Risk Decision Tree (SRDT; Cukrowicz et al., 2004; Joiner et al., 1999; Joiner et al., 2009) is a clinician-administered interview that addresses both desire and capability in determining suicide risk. Although several self-report instruments (Interpersonal Needs Questionnaire, INQ; and the Acquired Capability for Suicide Scale, ACSS) also examine these areas, the interview provides a more comprehensive assessment of the suicide risk framework and is appropriate when more time is available for suicide risk assessment. The SRDT interview also includes open-ended questions that allow the interviewer to probe for further information regarding individual items and investigates a wide range of risk factors, including those related to mental disorders. The SRDT interview examines suicide risk and suicidal desire. Questions investigate two components of desire: (1) lack of belonging, and (2) burdensomeness. The interview also reviews the capability for suicide, including suicidal plans and preparations, duration and intensity of suicidal ideation, history and number of past suicide attempts, means and opportunities, fearlessness of death, and recent stressful life events. This combined environmental and psychosocial information yields a suicide risk level. Low risk applies to people who have suicidal ideation but no plans or preparation and few other risk factors. Moderate risk is attributed to people who have multiple prior suicide attempts but no other current risk factors or "non-attempters" who have moderate to severe suicidal ideation and desire but no plans or preparation. High risk is reserved for people who have multiple suicide attempts or non-attempters who have multiple risk factors; high risk endorses both a plan and preparation for executing the plan (Joiner et al., 1999).

Availability and Cost

Although no formal SRDT instrument is publicly available, guidelines are available that describe how to administer the SRDT interview and include a visual representation of the decision tree matrix and sample items. The guidelines are available in the publication and at the web link listed below:

Cukrowicz, K. C., Wingate, L. R., Driscoll, K. A., & Joiner Jr, T. E. (2004). A standard of care for the assessment of suicide risk and associated treatment: The Florida State University Psychology Clinic as an example. *Journal of Contemporary Psychotherapy*, 34(1), 87–100. http://link.springer.com/article/10.1023/B:JOCP.0000010915.77490.71

Recommendations for Suicide Risk Screening Instruments

Information describing suicide screening instruments is based on a critical review of the existing literature. Key areas considered in making recommendations about suicide screens include empirical evidence supporting the reliability and validity of instruments, the relative costs of instruments, ease of administration, use within the criminal justice system, and alignment with theoretical frameworks that have been established for assessment of suicide risk. As noted previously, offenders who are screened as having significant suicide risk should be immediately referred for further assessment to determine the need for treatment, close supervision, and other services.

For brief suicide screening, the following instruments are recommended:

1. The Interpersonal Needs Questionnaire (INQ) coupled with the Acquired Capability for Suicide Scale (ACSS). The INQ/ACSS was developed based on the Suicide Risk Decision Tree and measures specific factors associated with suicide risk, including

suicidal desire (feelings of burdensomeness, lack of belonging) and capability.

(or)

2. The Beck Scale for Suicide Ideation (BSS).

(or)

3. The Adult Suicidal Ideation Questionnaire (ASIQ).

The BSS and ASIQ assess some, but not all components of the prevailing suicide risk assessment framework, but both instruments have been examined within the criminal justice system, and have been found to reliably predict suicide risk.

Each of the previously described instruments requires between 10–15 minutes to administer and score.

If additional time is available to provide a more detailed assessment of suicide risk, the following instrument is recommended:

- The Suicide Risk Decision Tree (SRDT), a clinician-administered interview that provides a comprehensive assessment of environmental and psychosocial factors associated with suicide risk. The SRDT examines factors that are fully aligned with the theoretical framework for suicide risk assessment, and its open-ended response format facilitates additional interviewer probes to follow up on specific questions.

The SRDT interview requires approximately 20 minutes to administer.

In contrasting the recommended suicide risk instruments, considerations should include the cost of these instruments. The BSS and ASIQ are commercially available and are more expensive to administer than the INQ/ACSS instruments, which are available in the public domain. However, the validity of the INQ/ACSS has not been determined within criminal justice settings. Although the Suicide Risk Decision Tree (SRDT) interview provides broader coverage of suicide risk factors, it requires additional time to administer.

Screening and Diagnostic Instruments for Trauma and PTSD

People with CODs have very high rates of trauma and posttraumatic stress disorder (PTSD) in comparison to the general population, and these rates are augmented in the criminal justice system (Elbogen et al., 2012; Lynch et al., 2013; Proctor, 2012; Proctor & Hoffmann, 2012; Steadman et al., 2013). Trauma is often overlooked in screening within the criminal justice system, particularly in substance use treatment settings. Failure to identify trauma within this population often leads to poor treatment outcomes (Prendergast, 2009; Ruiz, Douglas, Edens, Nikolova, & Lilienfeld, 2012; Steadman et al., 2013). Several specialized screening and assessment instruments have been developed to examine the history of trauma and PTSD, which may be useful within criminal justice settings. Several other general mental health screening and assessment instruments that also examine trauma and PTSD (e.g., CMHS, MINI, PAI, SCID-IV) are described in previous sections of this monograph. Screens for trauma and PTSD are generally brief, noninvasive, and do not require administration by a mental health professional. Two types of screening instruments are available: (1) those that address stressful life events and their effects, and (2) those that address severity of symptoms based on DSM criteria. The diagnostic screens are somewhat longer to administer but provide a formal diagnosis of PTSD and are often used as follow-ups to brief screens. As mentioned previously, screening for trauma/PTSD can be conducted by nonclinicians through use of standardized self-report instruments, which require minimal training. However, all staff who administer trauma screens should be fully aware of appropriate referral sources and the nature of trauma-related services. Offenders who screen positively as having significant problems related to trauma and PTSD should receive a thorough

assessment by a trained and licensed/certified mental health professional.

Changes to the DSM-5 Diagnostic Criteria for PTSD

There are several major differences between the DSM-IV criteria for PTSD and the more recent DSM-5 criteria (APA, 2013). The DSM-IV defined PTSD with the following criteria: A—traumatic event experienced, including severity, frequency, and intensity; B—re-experiencing traumatic events; C—avoidance of trauma; and D—hyperarousal. Criterion E assessed duration of traumatic symptoms and Criterion F assessed related functional impairment. Under DSM-5, PTSD is included in a new section, entitled, "Trauma and Stress-related Disorders." Criterion A now explicitly addresses sexual violation as a traumatic event and includes reoccurring exposure to traumatic events, such as those faced by law enforcement or paramedics. Moreover, Criterion A no longer requires a response of intense fear, helplessness, or horror. A new Criterion D ("negative cognitions and mood") has been added to capture symptoms related to distorted thinking and negative emotions. These symptoms were originally addressed in DSM-IV Criterion C. The new criterion includes items aimed at assessing persistent feelings of blame (self or others), detachment from others, anhedonia (inability to experience pleasure), and difficulty recalling traumatic events. Criterion E ("alterations in arousal") now examines changes in arousal and reactivity. Items include irritability and anger, reckless or impulsive behaviors, hypervigilance, difficulty sleeping, and difficulty concentrating. Criterion F has also been revised to describe the duration of symptoms, while the new Criterion G assesses functional impairment.

Screening Instruments for Trauma/PTSD

Impact of Events Scale–Revised (IES-R)

The IES is a 15-item self-report measure describing the current level of subjective stress experienced as a consequence of experiencing a traumatic event (Horowitz, Wilner, & Alvarez, 1979). The revised IES-R (Weiss, 2004; Weiss & Marmar, 1997) includes 22 items, with six additional items examining hyperarousal (e.g., exaggerated startle, psychophysiological arousal when reminded of the event) and one item that examines re-experiencing traumatic events. IES items are based on DSM-III-R/DSM-IV criteria. The three scales include avoidance, intrusion, and hyperarousal. Respondents indicate distress from zero (not at all) to four (extremely) on each item and questions inquire about symptoms experienced over the past 7 days. The cut-off score for the presence of PTSD is ≥33. Guidelines for scoring and interpretation are provided. The IES-R is one of the most widely used measures of PTSD symptoms. Unlike the majority of trauma/PTSD instruments, the IES-R addresses a wide range of traumatic experiences.

Positive Features

- The IES has adequate reliability and concurrent and discriminant validity, and has a cohesive factor structure (Creamer, Bell, & Failla, 2003)
- The IES is easy to administer and has been used with a variety of populations
- The IES has been used with offenders (Austin-Ketch et al., 2012)
- The IES-R uses a parallel format to that of the SCL-90-R, allowing for comparison of symptoms across instruments (Weiss, 2004)
- The IES-R can be used as an alternative to the PCL-C
- The IES-R is available in several languages, including Spanish (Báguena et al., 2001), Chinese (Wu & Chan, 2003), French (Brunet, St-Hilaire, Jehel, & King,

2003), German (Maercker & Schuetzwohl, 1998), and Japanese (Asukai et al., 2002)

- The IES-R has been used with veterans (Amdur & Liberzon, 2001; Forbes et al., 2003) and people with substance use disorders (Rash, Coffey, Baschnagel, Drobes, & Saladin, 2008; Schumacher, Coffey, & Stasiewicz, 2006)
- Among those who have substance use disorders with and without PTSD (Rash et al., 2008), the IES-R shows good diagnostic accuracy at a cut-off score of 33, as indicated by the Clinician Administered PTSD Scale (CAPS). The IES-R also has good overall accuracy (73 percent), sensitivity (73 percent), specificity (72 percent), positive predictive value (78 percent), and negative predictive value (67 percent). The IES-R demonstrates good convergent validity with the CAPS (r scores range .36–.60) and concurrent validity with the SCL-90-R (r scores range .47–.72) among people who have substance use disorders (Rash et al., 2008)
- The IES-R has good diagnostic accuracy among treatment-enrolled veterans who meet PTSD criteria (Creamer, Bell, & Failla, 2003), as indicated by the PTSD checklist (PCL; Weathers, Litz, Herman, Huska, & Keane, 1993), with an overall accuracy of 88 percent at a cut-off score of 33, sensitivity of 91 percent, specificity of 82 percent, positive predictive value of 90 percent, and negative predictive value of 84 percent. The IES-R and its subscales also have good convergent validity with the PCL within this same population (r scores range .70–.86; Creamer et al., 2003)
- In a large law enforcement sample, the IES-R and its subscales show good convergent validity with the Mississippi Scale for Combat-Related PTSD, Civilian Version (Keane, Caddell, & Taylor, 1988), with r scores ranging .53–.57 (Weiss & Marmar, 2004). The IES-R is also highly correlated with other measures of concurrent validity (r scores ranged .31–.50; Weiss & Marmar, 2004), including the Peritraumatic Dissociative Experiences Questionnaire (PDEQ, Marmar, Weiss, & Metzler, 1997), the Peritraumatic Distress Inventory (PDI, Brunet et al., 2001), and Depression and Global Symptom Index (GSI) scores on the SCL-90-R
- Factor analyses of the IES-R support a three-factor structure, in accordance with the three scales of avoidance, intrusion, and hyperarousal (Weiss & Marmar, 2004)
- Internal consistency of the IES-R is quite good across the three scales, including avoidance (alpha = .84), intrusion (alpha = .89), and hyperarousal (alpha = .82; Weiss & Marmar, 2004). Internal consistency across the IES-R scales is also quite good among veterans (alphas range .81–.87; Creamer et al., 2003) and people who have substance use disorders (alphas range .85–.91; Rash et al., 2008). Internal consistency of translated versions of the IES-R is also quite good (alphas range .83–.91; Weiss & Marmar, 2004)
- The test-retest reliability of the IES-R is quite good (r scores range .89–.94) over a 6-month period (Weiss & Marmar, 1996). Test-retest reliability of translated versions of the IES-R is also good (r scores range .52–.86; Weiss & Marmar, 2004)

Concerns

- Instructions must be provided to respondents for IES-R questions that ask about specific traumatic events
- The IES-R does not provide a diagnosis of PTSD and instead provides an evaluation of avoidance and intrusive symptoms
- The IES-R has not been widely studied among criminal justice populations
- At a cut-off score of 33, accuracy in determining the presence of PTSD may be low (kappa = .47; Rash et al., 2008)
- There has been inconsistent support for a three-factor structure of the IES-R, as several studies indicate one and two-factor structures (Báguena et al, 2001; Creamer et al., 2003; Taylor, Kuch, Koch, Crockett,

& Passey, 1998; Wagner & Waters, 2014). Other studies support a different three-factor structure (intrusion/hyperarousal, avoidance, and sleep/irritability/concentration; Asukai et al., 2002), or a four-factor structure (Amdur & Liberzon, 2001; King, Leskin, King, & Weathers, 1998). These findings suggest that the IES-R may measure general trauma-related distress rather than symptoms of PTSD

- Internal consistency of the IES-R is somewhat low across the three scales among veterans enrolled in treatment (alpha range .52–.83, Creamer et al., 2003)

Availability and Cost

The IES can be obtained at no cost at the following site: http://serene.me.uk/tests/ies-r.pdf

The instrument can also be found in the following articles: (1) Weiss, D. S., & Marmar, C. R. (1996). The impact of event scale–revised. In J. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD* (pp. 399–411). New York: Guilford. (2) Weiss, D. S., & Marmar, C. R. (2004). The impact of event scale–revised. In J. P. Wilson & T. M. Keane (Eds.), *Assessing psychological trauma and PTSD*, (2nd ed., pp. 168–189). New York: Guilford.

Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5)

The most recent version of the PCL, the Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5; Weathers et al., 2013), includes 20 items that examine the expanded DSM-5 PTSD criteria. The National Center for PTSD, operated by the U.S. Department of Veterans Affairs, recommends that the PCL-5 be administered in conjunction with the Life Events Checklist for DSM-5 (LEC-5) to obtain a more comprehensive measure of traumatic events experienced (Criterion A related to PTSD; VA, 2015). A severity score on the PCL-5 can be obtained by summing the scores for each of the 20 items. Preliminary recommendation by the National PTSD Center and the Department of Veterans Affairs suggests a cut-off score of 38 for determining PTSD diagnosis (Weathers et al., 2013). The previous version of this instrument included the PCL (Posttraumatic Stress Disorder Checklist), a 17-item self-report measure that is based on the DSM-IV criteria. The PCL is used to screen for PTSD symptoms, provide a diagnostic impression for PTSD, and monitor change in symptoms over time (Weathers et al., 1993).

Several versions of the previous PCL instrument (based on DSM-IV PTSD criteria) were designed for military (PCL-M) and civilian (PCL-C) populations. The PCL-M queries about symptoms related to traumatic military experiences and may be used with veterans or active service personnel. When considering which version to use, one should also take into account that individuals in the military may also have premilitary trauma experiences, and as such the PCL-C may also have utility for the veteran population. The PCL-C queries about symptoms related to traumatic life events and can be used with various populations. The PCL-Specific (PCL-S) queries about symptoms related to a specific traumatic life event. Symptoms identified by the PCL can refer to one or more traumas experienced. Prior to administering the PCL, it is important to screen respondents for Criterion A of DSM criteria for PTSD or the experience of an actual stressor involving actual or threatened death, serious injury to self or others, or actual or threatened sexual violence. The PCL requires approximately 10 minutes to administer. Respondents are asked to rate the severity of symptoms, according to "how much you have been bothered by the problem" during the past month, on a 1–5 scale. Total symptom severity is reflected in the summed score of the 17 PCL items. Thresholds for symptom severity include ratings of 3 or above on criterion B (re-experiencing symptoms, questions 1–5), 3 or above on Criterion C (avoidance of symptoms, questions 6–12), and 2 or above on Criterion D (hyperarousal, questions 13–17). Suggested cut-off scores for the PCL are 30–35 in community samples, 36–44 in medical clinics (e.g., VA primary care), and 45–50 in mental health

settings (Blanchard, Jones-Alexander, Buckley, & Forneris, 1996). The TCU Mental Trauma and PTSD Screen (TCU TRMAForm) is a version of the PCL used with offenders that is available from the Texas Christian University Institute of Behavioral Research.

Positive Features

- The PCL has been widely used with offenders (Ball, Karatzias, Mahoney, Ferguson, & Pate, 2013; Owens, Rogers, & Whitesell, 2011; Pankow et al., 2012; Rowan-Szal, Joe, Bartholomew, Pankow, & Simpson, 2012; Wolff, Frueh, Shi, & Schumann, 2012), including use to monitor change in PTSD symptoms while offenders are involved in treatment (Ball et al., 2013; Wolff et al., 2012)
- The PCL has been found to have greater diagnostic accuracy than several other screens (McDonald & Calhoun, 2010), including the four-item SPAN (startle, physically upset by reminders, anger, and numbness; Yeager, Magruder, Knapp, Nicholas, & Frueh, 2007) and the Primary Care PTSD Screen (PC-PTSD; Prins et al., 2003)
- The PCL can be used to monitor change in symptoms over time, particularly in treatment settings (McDonald & Calhoun, 2010)
- Across clinical, primary care, veteran, hospital, and community settings (McDonald, & Calhoun, 2010), the different versions of the PCL provide fair to good diagnostic accuracy at a cut-off score of 50, as determined by the CAPS, the SCID, and the MINI. However, other cut-off scores may be preferred based on the particular setting
- Among a military primary care sample (Gore et al., 2013), and using a cut-off score of 31, the PCL-C shows good diagnostic accuracy in comparison to the PTSD Symptom Scale Interview (PSS-I, Foa, Riggs, Dancu, & Rothbaum, 1993) at a cutoff of 31, with good sensitivity (93 percent), specificity (90 percent), and overall diagnostic accuracy (90 percent)
- Among women with substance use disorders (Harrington & Newman, 2007) and at a cut-off score of 44, the diagnostic accuracy of the PCL is better than the CAPS in identifying PTSD, with good overall accuracy (76 percent), sensitivity (76 percent), specificity (79 percent), positive predictive value (68 percent), and negative predictive value (80 percent)
- The concurrent validity of the PCL among female offenders was established in reference to the TCU Drug Screen (TCUDS), the TCU Psychological Functioning Scale, and the TCU social functioning scales (Rowan-Szal et al., 2012). Concurrent validity of the PCL was also established across measures of mental health and substance use among male offenders, individuals enrolled in community substance use treatment (Pankow et al.,2012), and parolees and probationers (Owens et al., 2011)
- Interrater reliability of the PCL is acceptable among community and clinical samples (Blanchard et al., 1996; Bollinger, Cuevas, Vielhauer, Morgan, & Keane, 2008; Keen, Kutter, Niles, & Krinsley, 2008) and veterans (Weathers et al., 1993)
- Internal consistency of the PCL and its scales is quite good among offenders (alphas range .73–.94 Rowan-Szal et al., 2012) and those who have severe mental disorders (.72–.87; Mueser et al., 2001)
- Confirmatory factor analysis indicates that the PCL has a three-factor structure, reflecting the three scales of re-experiencing, avoidance, and hyperarousal (Rowan-Szal et al., 2012)
- Test-retest reliability of the PCL-C is good over intervals of 1 hour (r score = .92), 1 week (r score = .87–88), and 2 weeks (r score = .68) among undergraduate students who had experienced a traumatic event (Adkins, Weathers, McDevitt-Murphy, & Daniels, 2008; Ruggiero, Del Ben,

Scotti, & Rabalais, 2003). The test-retest reliability of the PCL-M is quite good among military combat veterans, over a 1-week interval (r score = .96; Weathers et al., 1993)

Concerns

- Further study is needed to determine the diagnostic validity of the PCL among offenders
- The PCL does not assess all DSM criteria, including the types of traumatic event experienced, the duration of symptoms, negative cognitions, and clinical impairment related to daily functioning
- The PCL should not be used as the sole diagnostic instrument for PTSD, as it does not demonstrate the same diagnostic effectiveness as clinician-administered interviews (McDonald & Calhoun, 2010; National Center for Posttraumatic Stress Disorder, 2008), and further, it is geared toward DSM-IV
- PTSD symptoms often overlap with other mental health symptoms and thus can contribute to low rates of diagnostic accuracy (e.g., false positives) when using the PCL (McDonald & Calhoun, 2010)
- Various cut-off scores are recommended for different samples. Those administering the PCL should thus be aware of population base rates and specific cut-off scores for these populations
- The factor structure of the PCL-S may differ across settings, particularly because it references specific trauma rather than overall trauma history. Thus, scores on the PCL should be interpreted with caution, and interpretation should take into account the type of sample and related base rates for trauma history (Elhai et al., 2009)
- Interrater reliability of the PCL varies across samples. Particularly, low kappas (≤.50) have been found in primary care settings (Walker, Newman, Dobie, Ciechanowksi, & Katon, 2002; Yeager et al., 2007)

Availability & Cost

The PCL-5 can be obtained free of charge by completing an electronic request form, and information regarding changes from the previous PCL-C (based on the DSM-IV) to the newer PCL-5, including administration, scoring, and interpretation can be found at the following site: http://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp

The previous PCL instrument and all of its versions (e.g., PCL-C) can be downloaded at no cost at the following site: http://at-ease.dva.gov.au/professionals/assess-and-treat/ptsd/

The Life Events Checklist for DSM-5 (LEC-5) is a public domain instrument, and is available for download at the following site: http://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp

The TCU Mental Trauma and PTSD Screen (TCU TRMAForm) can be downloaded at no cost at the following site: http://ibr.tcu.edu/forms/client-%20health-and-social-risk-forms/

Primary Care PTSD Screen (PC-PTSD)

The PC-PTSD (Prins et al., 2003) is a four-item screen for PTSD in primary care settings. The PC-PTSD examines several symptoms of PTSD, including re-experiencing a traumatic event, emotional numbing, avoidance, and hyperarousal. Instructions query about traumatic experiences in the past month. The cut-off for indicating the presence of PTSD is a score of ≥ 3 positive responses. The PC-PTSD has variable cut-off scores, depending on the base rates of PTSD in different populations. Maximizing sensitivity over specificity is preferred in clinical settings in order to minimize false negatives, which can prove to be more costly in the diagnostic process (Calhoun et al., 2010). In using the PC-PTSD for screening of PTSD among those with CODs and in determining diagnoses, it is important to consider overlapping mental health and substance problems and their relationship with PTSD symptoms. People

screened as positive on the instrument should receive further clinician-administered assessment related to PTSD.

Positive Features

- The PC-PTSD is widely used in VA primary care settings (U.S. Department of Veterans Affairs [VA], 2004; VA/ Department of Defense, 2003)
- The PC-PTSD is designed for those with an eighth-grade reading level or higher
- The PC-PTSD has been used in various criminal justice settings (Ford, Chang, Levine, & Zhang, 2012; Ford & Trestman, 2005; Ford et al., 2007), including veteran treatment courts (Slattery et al., 2013)
- The Correctional Mental Health Screen (CMHS) has adapted items from the PC-PTSD (Ford & Trestman, 2005; Ford et al., 2007) to screen for PTSD in criminal justice settings
- Among those enrolled in substance use treatment, the PC-PTSD demonstrates acceptable sensitivity (67 percent) and specificity (72 percent) relative to a SCID-IV PTSD diagnosis (van Dam, Ehring, Vedel, & Emmelkamp, 2010)
- In primary care settings, as compared to the CAPS, the PC-PTSD shows good diagnostic accuracy at a cut-off score of 3, indicated by the AUC (92 percent), in addition to good sensitivity (85 percent), specificity (82 percent), and negative predictive value (98 percent; Freedy et al., 2010). Using a cut-off score of 3 in military primary care settings (Gore, Engel, Freed, Liu, & Armstrong, 2008), the PC-PTSD shows good sensitivity (70 percent), specificity (92 percent), and negative predictive value (97 percent) relative to the Posttraumatic Symptom Scale Interview (PSS-I, Foa et al., 1993)
- Among veterans, the PC-PTSD shows good sensitivity (83 percent), specificity (85 percent), and overall diagnostic accuracy (85 percent) at a cut-off score of 3, as determined by the SCID-IV for PTSD (Calhoun et al., 2010)
- At a cut-off score of 2 in a sample of veterans in primary care settings (Ouimette, Wade, Prins & Schohn, 2008), the PC-PTSD has higher specificity (96 percent) and overall diagnostic accuracy (93 percent) than the General Health Questionnaire (GHQ-12; Goldberg & Williams, 1988) and provides greater predictive validity than the GHQ in identifying PTSD
- Item response theory (IRT) analyses indicate that the PC-PTSD performs consistently well in screening for PTSD across gender groups (Oliver, 2013)
- The test-retest reliability of the PC-PTSD is quite good in primary care settings (r score = .83; Prins et al., 2003)

Concerns

- The PC-PTSD was designed for use in primary care settings and has not been widely studied in criminal justice settings
- The PC-PTSD does not identify specific traumatic life events related to PTSD symptoms (VA, 2013)

Availability and Cost

The PC-PTSD can be downloaded for free at the following site, which also provides instructions for administration and scoring of the instrument: http://www.ptsd.va.gov/PTSD/professional/pages/assessments/assessment-pdf/pc-ptsd-screen.pdf

Trauma Symptom Checklist (TSC-40)

The TSC-40 (Elliot & Briere, 1992) is a 40-item self-report measure of posttraumatic distress and associated symptoms related to events occurring throughout the lifespan. Respondents rate how often they have experienced each event on a four-point scale. The instrument contains six scales: anxiety, depression, dissociation, sexual abuse trauma index, sexual problems, and sleep disturbance. The TSC-40 is an improved version of the TSC-30 and includes items related to sexual

problems and sleep disturbance. The instrument is scored by summing each domain and/or by calculating a total score. Overall scores range 1–40. The recommended cut-off score for the presence of PTSD related traumatic stress is ≥ 23. The TSC-40 should not be used as a stand-alone instrument to identify PTSD but should rather be used in combination with a screening or assessment instrument for PTSD.

Positive Features

- The TSC-40 is a public domain instrument
- The TSC-40 is brief to administer
- The TSC-40 has been used with offenders, including those with CODs (Covington, Burke, Keaton, & Norcott, 2008; Grella, Stein & Greenwall, 2005; Hannah, Young & Moore, 2009; Messina & Grella, 2006; Messina et al., 2007; Zlotnick, Johnson, Najavits, 2009)
- Among female offenders, for every additional exposure to childhood traumatic events (as indicated by the LSC-R), the likelihood of a positive screen on the TSC-40 increases by 27 percent, supporting the concurrent and convergent validity of the TSC-40 (Messina & Grella, 2006)
- Among psychiatric inpatients, the total score of the TSC-40 correctly identifies 84 percent of individuals with sexual abuse, as determined by the Self-Rating Traumatic Stress Scale (SR-TSS; Davidson, Book, & Colket, 1995), supporting the concurrent validity of the instrument (Zlotnick et al., 1996). Used alone, the TSC-40 sexual abuse trauma index correctly identifies 77 percent of people who have a history of sexual abuse. Also supporting its concurrent validity, the TSC-40 scales of dissociation, anxiety, depression, and sexual abuse trauma index are moderately to strongly correlated with the SCL-90 scales of depression and anxiety, and the SR-TSS total scores (r scores range .40–.64)
- Among offenders, the concurrent validity of the TSC-40 is supported by findings that people with exposure to five or more traumatic events (as determined by the LSC-R) have higher mean scores on TSC-40 subscales (Messina et al., 2007)
- The concurrent validity of the TSC-40 among female drug court participants (Hannah et al., 2009) is supported by significant correlations between experiences of interpersonal abuse and child abuse, as determined by the LSC-R (r scores range .60–.61). In addition, 3-month follow-up scores on the TSC-40 for both anxiety and total score are significantly correlated with substance use (r scores range .50–.51)
- The TSC-40 can be used to monitor change in symptoms of PTSD during treatment (Zlotnick et al., 2009; Covington et al., 2008)
- The TSC-40 has good test-retest reliability, as demonstrated by significant correlations between baseline and 3-month follow-up scores across subscales (r scores range .50–.56)
- The TSC-40 total score has excellent internal consistency (Elliot & Briere, 1992; alpha = .90), as do the sleep disturbances (alpha = .77) and sexual problems (alpha = .73) scales. Other studies show similar results, with alphas ranging .66–.77 for the subscales; and alphas for the total score ranging .89–.91 (Briere, Elliott, Harris, & Cotman, 1995)

Concerns

- The psychometric properties of the TSC-40 have not been widely examined in criminal justice settings
- The TSC-40 was primarily designed for research purposes
- The TSC-40 may not be as comprehensive in scope as the TSI
- The TSC-40 does not examine traumatic life events that are experienced but rather associated posttraumatic distress and general psychological distress

Availability and Cost

The TSC-40 is a public domain instrument and can be downloaded at no cost at the following site, which also provides information regarding scoring and administration: http://bhpr.hrsa.gov/grants/areahealtheducationcenters/ta/Files percent20for percent20Veterans percent20Mental percent20Health percent20CE/traumachecklist.pdf

The Trauma Symptom Inventory (TSI)

The TSI is a 100-item self-report inventory that evaluates the presence of acute and chronic trauma symptoms. The instrument requires approximately 20 minutes to administer. The TSI contains 10 clinical scales that examine affective, cognitive, and physical issues related to trauma. Clinical scales include the following: Anxious Arousal (AA), Depression (D), Anger/Irritability (AI), Intrusive Experiences (IE), Defensive Avoidance (DA), Dissociation (DIS), Sexual Concerns (SC), Dysfunctional Sexual Behavior (DSB), Impaired Self-Reference (ISR), and Tension Reduction Behavior (TRB). Three validity scales are included to detect efforts to either underreport or exaggerate symptoms. These include Atypical Responses (ATR), Response Level (RL), and Inconsistent Responding (INC). Items are based on the DSM-IV symptom criteria for PTSD. Respondents rate the frequency of each symptom experienced on a four-point scale.

Separate TSI norms are available for men and women, as well as for different age groups. There is an 86-item alternative version (TSI-A) that does not examine sexual concerns or dysfunctional sexual behavior scales. A revised version of the TSI is also available (TSI-2; Briere, 2010), which provides improved validity scales for detecting malingering or feigned PTSD symptoms. The TSI-2 contains 136 items, two validity scales, 12 clinical scales, 12 subscales, and four factors. The TSI-2 was normed on a large U.S. sample. Additional clinical scales include Insecure Attachment (IA), Somatic Preoccupations (SOM), and Suicidality (SUI). The instrument provides a reliable index of change in symptoms over time. An alternate version is also available for the TSI-2 (the TSI-2A).

Positive Features

- The TSI is easy to administer and has been used extensively in a variety of clinical settings
- A survey of members of the International Society for Traumatic Stress Studies (ISTSS) indicates that the TSI is one of the most widely used self-report instruments for PTSD (Elhai, Gray, Kashdan, & Franklin, 2005)
- Computerized scoring of the instrument is available
- The TSI has been used with offenders (Bradley & Follingstad, 2003; Day et al., 2008; Goldenson, Geffner, Foster, & Clipson, 2007) and substance-involved populations (Adams et al, 2011; Najavits, & Walsh, 2012)
- The TSI contains three validity scales designed to detect the level, typicality, and consistency of responses (Briere, 1995)
- The ATR validity scale is effective in detecting feigned PTSD symptoms across race/ethnicity groups (Briere, 2010)
- Scores on the sexual concerns scale of the TSI are correlated with longer stay in substance use treatment among women (Adams et al., 2011)
- In a community sample of people (McDevitt-Murphy, Weather, & Adkins, 2005) reporting a traumatic event, TSI clinical scales are moderately to strongly correlated with relevant cluster symptoms of the CAPS. For example, the Intrusive Experiences scale is correlated with Cluster B symptoms of re-experiencing trauma on the CAPS (r score = .59). The TSI clinical scales also are positively correlated with other measures of convergent validity, including the IES- R (r scores range .36–.68), the PCL (r scores range .32–.65), the Civilian Mississippi Scale (CMS; r scores range .36– 66), and the Anxiety-

Related Disorders Scale (ARD-T) on the Personality Assessment Inventory (PAI, r scores range .35–.73). This same study found that the TSI demonstrates good diagnostic accuracy across subscales, as determined by the CAPS, with sensitivity ranging 63–94 percent and specificity ranging 59–91 percent. Cut-off scores were as follows: Defense Avoidance (T ≥62), Anxious Arousal (T ≥ 63), Depression (T ≥ 58), Atypical Response (T ≥ 52), and Intrusive Experiences (T ≥ 51)

- Among undergraduates instructed to feign PTSD symptoms, the Atypical Response Scale was able to accurately detect malingering as determined by the Personality Assessment Inventory (PAI) Negative Impression Management scale (NIM). At a cut-off score of 7, the TSI ATR scale accurately classifies 74 percent (sensitivity) of malingerers, and 77 percent (specificity) of those experiencing "true" PTSD distress, with an overall correct classification rate of 75 percent (Briere, 2010)
- The internal consistency of the TSI across subscales is quite good (alphas range .84–.97) in community, clinical, and domestic violence samples (Kaysen et al., 2007), among undergraduate students (Burns, Jackson, & Harding, 2010), and in military samples (Briere, 1995)
- The TSI has good internal consistency (alphas range .74–.90) and good sensitivity (91 percent) and specificity (92 percent; Briere, 1995)

Concerns

- Psychometric properties of the TSI have not been established in criminal justice settings
- The TSI is not a public domain instrument and is somewhat costly
- Advanced clinical training is recommended for staff assigned to interpret TSI test results
- Information is not available regarding test-retest reliability of the TSI scales

Availability and Cost

The TSI instrument is commercially available from the Psychological Assessment Resources, Inc., P.O. Box 998, Odessa, FL 33556; (800) 331-8378.

The TSI-2 can be purchased online at the following site: http://www4.parinc.com/products/Product.aspx?ProductID=TSI-2

The TSI introductory kit is relatively costly ($205) and contains the professional manual, 10 reusable item booklets, 25 hand-scorable answer sheets, and 25 profile forms. Computerized software that includes scoring is relatively costly, at $355.

Screening Instruments for Traumatic Life Events and Associated Symptoms

Life Stressor Checklist (LSC-R)

The LSC-R (Wolfe & Kimerling, 1997) is a self-report measure that assesses stressful life events. The LSC-R contains 30 items that query about exposure to traumatic events, including natural disasters; accidents; physical/sexual abuse; and other stressful life events, such as divorce, foster care, and financial difficulties. Some events, like sexual abuse, are queried for occurrence in both childhood and adulthood. The instrument also includes an item specific to women (occurrence of abortion). For each item, respondents are asked to provide their age at the time of the event, and as relevant, the presence of a threat or serious injury to self/others, fear/helplessness experienced, and duration of distress. Respondents are asked to indicate up to three events that have caused the most impairment. Individuals who endorse traumatic events should be further assessed to determine the presence of PTSD.

Positive Features

- The LSC-R is brief to administer
- The LSC-R includes information specific to trauma experienced by women

- The LSC-R explicitly measures criterion A2 of the DSM-IV (experience of helplessness or horror)
- The LSC-R has been used in criminal justice settings (Grella, Stein, & Greenwall, 2005; Hannah et al., 2009; Messina & Grella, 2006; Messina et al., 2007; Wolff et al., 2011)
- The LSC-R has been used with law enforcement (Inslicht et al., 2010; Maguen et al., 2009; McCaslin et al., 2006), people with substance use disorders (Hannah et al., 2009; Harrington & Newman, 2007; Ouimette, Read, & Brown, 2005; Stewart, Grant, Ouimette, & Brown, 2006; Toussaint, VanDeMark, Bornemann, & Graeber, 2007), and those with CODs (Brown & Melchior, 2008; Giard et al., 2005)
- Among offenders, the LSC-R's concurrent validity is supported by significant correlations with different types of traumatic events, including physical abuse, sexual abuse, violence, and incarceration of a family member (Messina et al., 2007). Support for the concurrent validity of the LSC-R is also found among sex offenders, whose risk for sexual offending is predicted by experiences of sexual abuse, physical neglect, emotional abuse, and family violence (Jennings, Zgoba, Maschi, & Reingle, 2013)
- Female offenders with a history of conduct disorders, substance use treatment, and homelessness have greater exposure to traumatic events in childhood, as determined by the LSC-R, supporting the concurrent validity of the instrument (Messina & Grella, 2006). Female offenders experiencing childhood traumatic events (e.g., death of a family member, assault, accident), as determined by the LSC-R, also have a higher incidence of violent criminal behavior (Grella, Stein, & Greenwall, 2005)
- The concurrent validity of the instrument is also supported by findings that female drug court participants who have experienced child abuse, as identified by the LSC-R, are more likely to have alcohol or drug use disorders (Hannah et al., 2009). Additionally, female offenders who have mental disorders have significantly higher rates of exposure to traumatic life events, as identified by the LSC-R, particularly those who have experienced sexual abuse (Wolff et al., 2011)
- Among females who have CODs, the LSC-R has acceptable to excellent test-retest reliability over a 1-week interval across different types of events (kappas range .52–.97; McHugo et al., 2005)
- The interrater reliability of the LSC-R is quite good, as indicated by high agreement (79–98 percent) across endorsed events among females who have CODs (McHugo et al., 2005)

Concerns

- The psychometric properties of the LSC-R have not been established in criminal justice settings
- The ability of the LSC-R to predict PTSD has not been widely studied
- The LSC-R describes other stressful life events that may not meet Criterion DSM-IV A1 for PTSD

Availability and Cost

The LSC-R is a public domain instrument and can be downloaded without charge at the following site: http://www.ptsd.va.gov/PTSD/professional/assessment/te-measures/lsc-r.asp

Stressful Life Events Screening Questionnaire-Revised (SLESQ-R)

The SLESQ-R (Goodman, Corcoran, Turner, Yuan, & Green, 1998) is a 13-item self-report questionnaire that measures lifetime exposure to traumatic life events. The SLESQ-R was developed as a screening tool for potential PTSD. The stressful life events are those considered traumatic by Criterion A1 in the DSM-IV. The instrument includes 11 questions that examine

specific events experienced and 2 general questions that assess any other traumatic life events. Questions review experiences of physical/sexual abuse, military trauma, threatened death or injury to self or others, and actual death or injury to others. Respondents indicate whether the particular event occurred, the age at which the event occurred, frequency and duration of the event, and hospitalization or other consequences related to the event. Endorsement of a traumatic event should be followed by a formal assessment of PTSD symptoms.

Positive Features

- The SLESQ-R is brief to administer
- The SLESQ-R is available in Spanish
- Among people who have severe mental disorders, use of the SLESQ-R is recommended prior to administration of the PCL
- The SLESQ accurately identifies a range of traumatic life events experienced by low-income minority respondents (Green, Chung, Daroowalla, Kaltman, & DeBenedictis, 2006)
- Among undergraduate students, those with multiple traumatic life events identified by the SLESQ endorse higher trauma-related stress, as determined by the Traumatic Symptom Inventory (Green, Goodman et al., 2000)
- The reliability of the self-report and interview-administered versions of the SLESQ among undergraduate students is quite good across different traumatic life events (mean kappa = .77; median kappa = .64; Goodman et al., 1998)
- The test-retest reliability of the SLESQ over a 2-week interval is quite good among undergraduate students (r score = .89; Goodman et al., 1998)

Concerns

- The psychometric properties of the SLESQ-R have not been widely studied in criminal justice settings
- The SLESQ-R should not be used as a stand-alone instrument to identify PTSD, and those who endorse a traumatic event should receive a more comprehensive assessment for PTSD and trauma by a trained clinician.
- Respondents may report the same incident for multiple SLESQ-R questions, leading to inflation of scores. Thus, those administering the instrument should follow-up and record responses in the most appropriate category.
- The SLESQ-R only assesses criterion A1 of PTSD (experience of a traumatic life event) and does not query about other PTSD criteria
- The SLESQ-R may not provide broad coverage of all traumatic events included in criterion A1, thus potentially under-identifying those with PTSD symptoms (Long et al., 2008)
- Estimates of reliability and validity of the SLESQ-R were established with undergraduate students and not with diverse populations
- There may be differences in the reliability of reported traumatic events on the self-report and interview versions of the SLESQ. Specifically, under-reporting of events such as experienced child sexual/physical abuse may occur on the self-report version of the instrument (Green et al., 1998)
- The SLESQ can misidentify "true" traumatic events among low-income minority respondents (Green et al., 2006). For example, robbery, being threatened with a weapon, and attempted rape are sometimes identified by the SLESQ as stressors rather than as "true" traumatic events. However, miscarriage, abortion, emotional abuse, substance use, and eating disorders are sometimes identified as "true" traumatic events experienced but are not classified as traumatic events by the SLESQ. Therefore the SLESQ may not accurately identify "true" traumatic

events experienced by minorities, leading to potential under-diagnosis of PTSD

- Test-retest reliability in undergraduate students may be lower for life threatening events, attempted sexual assault, and "other" traumatic events, as indicated by kappas lower than .60 (Green et al., 1998)

Availability and Cost

The SLESQ-R is a public domain instrument and can be downloaded without charge at the following site: http://ctc.georgetown.edu/toolkit Direct link to the SLESQ-R form: https://georgetown.app.box.com/s/nzprmm2bn5pwzdw1l62w

Alternatively, the measure can be requested by e-mailing the developer of the measure, Dr. Lisa A. Goodman, at goodmalc@bc.edu

Information describing the SLESQ-R can be found at the following site: http://www.ptsd.va.gov/professional/assessment/te-measures/stress-life-events.asp

Trauma History Questionnaire (THQ)

The THQ (Green, 1996) is a 24-item self-report measure that examines traumatic events within different categories. The categories include crime-related events (items 1–4, e.g., robbery), general disaster (items 5–17, e.g., accidents involving injury to self/death of others, military trauma, natural disaster), and physical/sexual experiences (items 18–24, e.g., physical attacks, sexual assaults). Respondents are asked to indicate if they were exposed to the event, if it occurred repeatedly, the age at which it occurred, and the frequency of the event. The THQ requires approximately 10–15 minutes to complete. The THQ can be provided in an interview and requires approximately 15–20 minutes to administer. Positive endorsement of items should be followed up with a more formal assessment of PTSD symptoms.

Positive Features

- The THQ is brief to administer
- The THQ is designed for both clinical and research settings
- The THQ is available in Spanish
- The THQ has been used with offenders, including those who have substance use disorders and CODs (Komarovskaya, Booker-Loper, Warren, & Jackson, 2011; Lynch, Fritch, & Heath, 2012; Sacks, Sacks, McKendrick et al., 2008; Sacks, McKendrick, Sacks, Banks, & Harle, 2008; Sacks, McKendrick, Hamilton et al., 2008; Salgado, Quinlan, & Zlotnick, 2007; Sarkar, Mezey, Cohen, Singh, & Olumoroti, 2005)
- The THQ has been used among people who have severe mental disorders (Lommen & Restifo, 2009; Kilcommons & Morrison, 2005; Mueser et al., 2008, Mueser et al., 2007)
- Offenders who receive psychiatric services have higher rates of traumatic events on the THQ, particularly for physical and sexual abuse, in comparison to non-offender psychiatric patients (Sarkar et al., 2005)
- One study of the THQ found that all offenders were exposed to at least one traumatic event prior to committing an offense (Payne, Watt, Rogers, & McMurran, 2008)
- Female offenders determined by the THQ to have been exposed to interpersonal violence show significant levels of PTSD symptoms, as indicated by the PCL; general psychiatric distress, as indicated by the BSI; and recent substance use. Repeated interpersonal violence identified by the THQ predicts PTSD symptoms and general psychiatric distress (Lynch et al., 2012)
- According to the THQ, female offenders with polysubstance use disorders report higher rates of exposure to trauma in comparison to people with single types of substance use problems, supporting the concurrent validity of the instrument (Salgado et al., 2007)

- The convergent validity of the THQ with the SLESQ is quite good, with kappas for individual items ranging .61–1.00 in a large sample of depressed low-income women (Goodman et al., 1998). Similarly, the THQ exhibits significant correlations with a measure of exposure to conflict, the Conflict Tactics Scale (r score = .46), in a sample of battered women (Humphreys, Lee, Neylan, & Marmar, 1999)
- Supporting the predictive validity of the instrument among inpatient and outpatients who have severe mental disorders, the frequency of trauma events identified by the THQ predicts PTSD symptoms, as determined by the PCL (Mueser et al., 1998). In a law enforcement sample, the THQ contributes unique variance in predicting PTSD symptoms (Lilly, Pole, Best, Metzler, & Marmar, 2009). Other studies also show that the THQ is related to PTSD symptoms (Golier et al., 2003; Green, Krupnick et al., 2000; Najavits et al., 1998; Spertus, Yehuda, Wong, Halligan, & Seremetis, 2003) and depression (Spertus, Burns, Glenn, Lofland, & McCracken, 1999, Spertus et al., 2003)
- Test-retest reliability of the THQ over a 2-week interval ranges from acceptable to excellent (kappas = .57–.82; Mueser et al., 2001) across traumatic events reported by psychiatric inpatients. Similarly, interrater reliability is quite good (kappas = .76–1.00) across reported traumatic events (Mueser et al., 2001)
- Test-retest reliability of the THQ among college students is adequate over a 2–3 month period (r scores range .51–.90) across events (Green, Goodman et al., 2000; Green et al., 2005)

Concerns

- As with other trauma screens, the THQ should not be used as a stand-alone instrument in diagnosing PTSD and rather should be used in combination with other instruments that examine symptom severity
- It may be difficult to identify traumatic events as defined by PTSD Criterion A, as the THQ does not explicitly examine the newly revised DSM-5 PTSD Criterion A
- Respondents may underreport, overreport, or distort traumatic events, contributing to lower validity and reliability of the measure (Hooper, Stockton, Krupnick, & Green, 2011)
- The reliability of the THQ can be compromised during repeated administrations if the respondent reports the same traumatic event under a different category (Hooper et al., 2011)
- Test-retest reliability of the THQ for general events (e.g., other serious injury or other unwanted sexual incident) may be somewhat low (r score = .47; Hooper et al., 2011)

Availability and Cost

The THQ is a public domain instrument and can be downloaded at no cost at the following site: http://ctc.georgetown.edu/toolkit. Direct link to the THQ: https://georgetown.app.box.com/s/9ol8x4rwz8jgwo1bwgo8

Paper copies of the instrument can be obtained by sending a written request to the address listed below:

Bonnie L. Green, Ph.D.
Department of Psychiatry
Georgetown University
611 Kober Cogan Hall
Washington, DC 20007

The Trauma History Screen (THS)

The THS (Carlson et al., 2011) is a brief 13-item self-report measure that examines lifetime traumatic events experienced. The measure inquires about exposure to 11 specific events (e.g., military trauma, accident, natural disaster, physical/sexual abuse) and general events (any other threatening event). For each positively endorsed event, the respondent indicates the

number of times the event occurred. The total number of events identified provides an index of high magnitude stressors (HMS). A follow-up screening question asks if any of the positively endorsed event(s) causes significant distress. The total number of events endorsed as causing distress reflects the number of traumatic stressors (TS).

For events that are causing distress, the respondent is asked to complete information regarding the age at which the event occurred; a description of the event; if the event represented a threat that could lead to death or injury; and if there were feelings of helplessness, horror, and/or dissociation experienced because of the event. The THS also examines the duration of distress ("not at all" to "a month or more") and uses a five-point scale to measure the amount of distress experienced ("not at all" to "very much"). The THS is based on DSM-IV PTSD criteria and reviews persistent posttraumatic events (PPD) by describing the number of events that involved actual/threatened death or injury (Criterion A1 related to PTSD); experiences of fear, helplessness, or horror (Criterion A2); duration of distress of 1 month or more (Criterion E); and severity of distress. This information can be used to provide a diagnostic impression related to PTSD, but should be followed-up by use of a formal diagnostic instrument. The THS requires less than 10 minutes to complete.

Positive Features

- The THS can be used in both clinical, nonclinical, and research settings
- The THS requires only a sixth-grade reading level
- The THS is brief to administer
- The THS explicitly assesses DSM-IV Criterion A2 for PTSD (intense fear, helplessness/horror)
- The THS has been used in a variety of populations, including people with severe mental disorders (Zimbrón et al., 2013), college students who endorse at least one heavy drinking episode (Monahan et al., 2013; Murphy et al., 2012), active duty and military veterans (Carlson et al., 2011; Fanning & Pietrzak, 2013; Stein et al., 2012), and community samples (Carlson, Smith, & Dalenberg, 2013)
- The convergent validity of the THS high magnitude stressors (HMS) and persistent posttraumatic distress (PPD) is quite good among a sample of veterans who are homeless and have high rates of mental disorders (Carlson et al., 2011), as evidenced by strong correlations with trauma indicated by military records (r scores range .57–.87)
- The THS (Carlson et al., 2011) is highly correlated with another validated measure of stressful life events, the Traumatic Life Events Questionnaire (TLEQ), for reported HMS (r score = .77) and is also correlated with the PCL-C for reported HMS and PPD among veterans who are homeless (r scores range .25–.41), hospital trauma patients (r scores range .33–.38), university students (r scores range .18–.22), other young adults (r scores range .30–.34), and adults (r scores range .32–.37)
- Interrater reliability of the THS on HMS and PPD is quite good among veterans who are homeless (kappas = .70, .75, respectively), hospital trauma patients (kappa = .61, HMS only), university students (kappa = .74, HMS only), and young adults (kappa = .74, HMS only; Carlson et al., 2011)
- The test-retest reliability of HMS and PPD is high over a 1-week interval among veterans who are homeless (r scores range .73–.93), hospital trauma patients (.74–.95), university students (.82–.87), and other young adults (.73–.77; Carlson et al., 2011)

Concerns

- The THS has not been studied in criminal justice settings
- The THS is a fairly new measure and requires further research to determine relevant psychometric properties

- Scoring rules for the THS must be obtained from the original development paper (Carlson et al., 2011)
- The THS has more global items than other trauma instruments and could result in high "false negatives" because it may not accurately assess all traumatic stressors. Conversely, the instrument may produce high rates of "false positives" because it does not define the interval of persistent distress (Carlson et al., 2011)

Availability and Cost

The THS is a public domain instrument and can be downloaded without cost at the following site: http://www.midss.ie/sites/www.midss.ie/files/trauma_history_screen.pdf

Information describing the THS and paper forms of the instrument can be obtained at the following site: http://www.ptsd.va.gov/professional/assessment/te-measures/ths.asp

Diagnostic Instruments for PTSD

The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5)

The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) is a 30-item structured, clinician-administered interview that assesses PTSD diagnostic criteria for DSM-5 (CAPS-5; Weathers et al., 2013). The CAPS-5 is a structured interview that includes standardized questions and probes examining 20 PTSD symptoms, as reflected in revisions to the DSM-5 criteria that were described previously in this section. The instrument was developed to enhance the validity and reliability of PTSD diagnoses (Blake et al., 1995) by rating the frequency and intensity of each of the diagnostic symptoms of PTSD. Three versions of the CAPS-5 are available to assess for PTSD symptoms occurring in the past week, the past month, and over the lifetime. There is also a version for children and adolescents (CAPS-CA) that is being revised for DSM-5. The instrument can also be used to monitor changes in symptoms over the course of treatment and provides a more comprehensive and valid approach for diagnosing PTSD than use of brief screening instruments. The psychometric properties presented below under positive features and concerns are based on the prior DSM-IV version of the CAPS.

Major changes to the CAPS-5 include that the respondent report of PTSD symptoms is based on only one indexed traumatic life event, and each symptom is rated with a single severity score, on a scale from 0 ("Absent") to 4 ("Extreme/incapacitating") that accounts for both frequency and intensity of symptoms. A diagnosis of PTSD is made if an individual endorses moderate or higher severity (≥ 2) symptoms for at least one item from Criterion B, one item from Criterion C, two items from Criterion D, and two items from Criterion E. The disturbance, as in DSM-IV, should last at least 1 month and cause significant distress or impairment. Symptom cluster severity scores are generated by summing severity scores for items corresponding to a particular DSM-5 cluster. It is recommended that questions related to Criterion A are supplemented by administration of the Life Events Checklist for DSM-5 (LEC-5), which examines lifetime exposure to 16 events, and any other event that may potentially cause trauma and PTSD. The CAPS requires 45–60 minutes to administer. Scoring and interpretation guidelines are included in the CAPS-5.

As mentioned previously, instructions for the CAPS-5 recommend administering the LEC-5 (or another structured screen that reviews past traumatic life events) in advance of inquiring about specific events that might be related to PTSD. The LEC-5 is a 17-item instrument that can be administered via self-report or interview. An extended self-report version is available to identify the "worst" event (if there was more than one) that occurred during the designated time period. The interview version of the LEC-5 provides this same information, and helps to determine whether Criterion A for PTSD has been met.

Positive Features

- The CAPS is considered to be the "gold standard" for diagnosing PTSD
- The CAPS assesses current and past symptoms of PTSD
- The CAPS provides explicit anchors and behavioral referents to guide ratings
- In forensic settings, the CAPS is recommended for assessment of PTSD symptoms and diagnosis (Huang, Zhang, Momartin, Cao, & Zhao, 2006; Keane, Buckley, & Miller, 2003; Zlotnick, Najavits, Rohsenow, & Johnson, 2003; Zlotnick et al., 2009)
- The CAPS has been translated into Bosnian, Chinese, French, German, and Swedish
- The instrument has been used with diverse populations, including people who have mental and substance use disorders
- The CAPS has been used with offenders (Spitzer et al., 2001; Trestman, Ford, Zhang, & Wiesbrock, 2007)
- The CAPS has demonstrated excellent psychometric properties (convergent, discriminant, diagnostic validity, and sensitivity to clinical change) among clinical and research populations (Weathers, Keane, & Davidson, 2001)
- Relevant scales of the PCL are highly correlated with the CAPS (r scores range .58–.74), supporting the convergent validity of the CAPS (Palmieri, Weathers, Difede, & King, 2007). Additionally, in support of the concurrent validity of the CAPS, PTSD severity among veterans is higher for those with a history of arrest, depression, and substance use (Calhoun, Malesky, Bosworth, & Beckham, 2005)
- In clinical and nonclinical samples, the CAPS demonstrates high agreement with the Posttraumatic Stress Scale-Interview (PSS-I) for diagnosis of PTSD, when employing scoring rules defined by Blanchard et al. (1995; kappas ≥ .55; Foa & Tolin, 2000). The CAPS also demonstrates high correlations between its subscales and the PSS-I (Foa & Tolin, 2000)
- Intraclass correlations with the CAPS total score is quite good among people who have severe mental disorders, (.97; Mueser et al., 2008), as are correlations across each criterion (ICCs range .91–.99; Mueser et al., 2001)
- Interrater reliability for a PTSD diagnosis is quite good among samples of people who have severe mental disorders (kappas range .91–1.0; Mueser et al., 2001; Mueser et al., 2008)
- Interrater reliability among veterans is quite good for a categorical diagnosis of PTSD (kappa = .92; Calhoun et al., 2005)
- Interrater reliability (Hovens et al., 1994) is high across frequency (kappas range .92–1.00), intensity ratings (kappas range .93–.98), and global severity ratings (r score =.89)
- Internal consistency is quite good for frequency (alphas range .63–.85), intensity (alphas range .71–.81), and total score (alpha = .94; Mueser et al., 2001) among people who have severe mental disorders. Similar results were found among clinical and nonclinical samples, with alphas ranging .71–.88 (Foa & Tolin, 2000)
- Test-retest reliability of the CAPS over a 2-week interval among people with severe mental disorders is acceptable (kappa = .63; Mueser et al., 2001) and at a severity score of ≥ 65, reliability is higher (kappa = .90)

Concerns

- The CAPS is quite lengthy to administer
- A significant amount of training is required to conduct CAPS interviews
- The CAPS is designed for research purposes and may not be ideally suited for routine use in clinical settings
- The psychometric properties of the CAPS have not been widely studied in criminal justice settings

- The intensity ratings for individual PTSD symptoms may be difficult to ascertain from the range of symptoms identified
- Scoring rules for diagnosis of PTSD using the CAPS may vary by definition (see Blanchard et al., 1995; Weathers, Ruscio, & Keane, 1999), and liberal versus stringent scoring criteria can result in different rates of PTSD diagnosis, and inconsistencies in diagnostic agreement between the CAPS and other interview measures of PTSD (PSS-I; Foa & Tolin, 2000)

Availability and Cost

The Clinician-Administered PTSD Scale for DSM-5 (CAPS-5) is a public domain instrument that can be obtained at no cost via an online request form at the following site: http://www.ptsd.va.gov/professional/assessment/adult-int/caps.asp

Information regarding scoring of the CAPS-5 is available at the same website. In the past, a CAPS training manual and a CAPS training CD could be obtained from the National Center for PTSD, operated by the U.S. Department of Veterans Affairs.

The Life Events Checklist for DSM-5 (LEC-5) is a public domain instrument and is available for download at the following site: http://www.ptsd.va.gov/professional/assessment/te-measures/life_events_checklist.asp

Posttraumatic Stress Diagnostic Scale (PDS)

The PDS (Foa, 1996) is a 49-item self-report measure that assesses severity (Criterion B, C, and D) of PTSD symptoms related to a traumatic event. Items assess all DSM-IV criteria for PTSD. Current (past month) PTSD is addressed and instructions can be adapted for other time frames (e.g., lifetime). The PDS addresses traumatic events experienced (Criterion A), duration of symptoms (Criterion E), and functional impairment (Criterion F). There are four sections of the PDS, including (1) a trauma checklist; (2) description of traumatic events provided by the respondent, with queries for injuries, serious threats to life, helplessness, and terror; (3) assessment of 17 DSM-IV PTSD symptoms; and 4) functional impairment. Total severity scores on the PDS range 0–51. The recommended cut-off score for diagnosis of PTSD is ≥ 27. A profile report can be generated that reviews PTSD diagnosis, symptom frequency, symptom severity, and level of functional impairment. The PDS can be used for screening of PTSD symptoms and for diagnosis of PTSD.

Positive Features

- The PDS is highly recommended for assessment of PTSD symptoms (Keane, Silberbogen, & Weierich, 2008)
- The PDS is a commonly used tool among the International Society for Traumatic Stress Studies (ISTSS; Elhai et al., 2005)
- The PDS has been used with offenders (Harner, Budescu, Gillihan, Riley, & Foa 2013; Messina, Grella, Cartier, & Torres 2010; Sacks et al., 2008)
- Concurrent validity of the PDS is quite good (Foa, Cashman, Jaycox & Perry, 1997), as demonstrated by strong correlations with the State-Trait Anxiety Inventory (STAI) and the IES-R
- The PDS demonstrates good diagnostic accuracy, with overall accuracy ranging 82–88 percent. At a cut-off score of 27, the PDS also has acceptable sensitivity (67–89 percent), specificity (75–91 percent), and negative predictive value (86–96 percent) among psychiatric outpatients and those seeking treatment for PTSD, in addition to those who are at high risk for trauma (Foa et al., 1997; Sheeran & Zimmerman, 2002)
- Among sexual assault survivors, drinking problems to cope with PTSD symptoms is a significant predictor of severity scores on the PDS (Ullman, Filipias, Townsend, & Starzynski, 2006). Moreover, severity scores on the PDS are significantly correlated with alcohol problems as

measured by the MAST (Ullman, Filipias, Townsend, & Starzynski, 2005)

- The PDS shows high internal consistency across domains (alphas range .78–.92; Foa et al., 1997)
- Test-retest reliability of the PDS is quite good for diagnosis (kappa = .74) and severity scores (r scores range .77–.85) among those endorsing a traumatic experience (Foa et al., 1997)

Concerns

- The PDS has not been extensively studied in adult criminal justice settings
- The PDS may overdiagnose PTSD, as indicated by high rates of "false positives" among a sample of domestic violence survivors (Griffin, Uhlmansiek, Resick, & Mechanic, 2004). Thus, caution should be exercised when interpreting PDS scores among certain populations (Keane et al., 2008)
- The PDS is highly correlated with the BDI, and as such, the instrument may not provide adequate discriminant validity in distinguishing between depressive symptoms and PTSD (Foa et al., 1997; Norris & Hamblen, 2004)
- The self-report nature of the PDS may detract from its validity in diagnosing PTSD

Availability and Cost

The PSD has been updated to the PDS-5 for the *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*. To obtain the PDS-5 with information about its administration and use, please contact Ellen Kubis at the University of Pennsylvania's Center for the Treatment and Study of Anxiety at ekubis@pennmedicine.upenn.edu.

Posttraumatic Symptom Scale–Interview Version (PSS-I)

The PSS-I is a semi-structured interview that provides both diagnosis of PTSD and assessment of PTSD symptom severity. The PSS-I includes 17 items that assess DSM-IV PTSD symptoms related to re-experiencing (items 1–5), avoidance (items 6–12), and hyperarousal (items 13–17). Items inquire about frequency and severity. Scoring is calculated by summing items within each domain, and a total score is obtained by summing all 17 items across domains. A diagnosis is made based on achieving a score of "2" or more in each domain. The PSS-I asks about current PTSD symptoms (past month or past 2 weeks). The PSS-I requires approximately 15–25 minutes to administer.

Positive Features

- The PSS-I is a brief semi-structured interview that performs as well as the CAPS in assessing PTSD and is briefer to administer (International Society for Traumatic Stress Studies, 2013)
- The PSS-I has been used successfully among people who have severe mental disorders (Brunet, Birchwood, Upthegrove, Michail, & Ross, 2012; O'Hare, Sherrer, & Shen, 2006), offenders (Sacks, McKendrick, & Hamilton, 2012), people with substance use problems (Foa & Williams, 2010; Reynolds et al., 2005), those with co-occurring PTSD and substance use disorders (Foa & Williams, 2010; Riggs, Rukstalis, Volpicelli, Kalmanson, & Foa, 2003), and in community samples (Bedard-Gilligan, Jaeger, Echiverri-Cohen, & Zoellner, 2012; O'Hare, Sherrer, Yeamen & Cutler, 2009)
- The diagnostic accuracy of the PSS-I is quite good in clinical and nonclinical samples (Foa & Tolin, 2000), with sensitivity ranging 71–86 percent and specificity ranging 78–100 percent for different scoring approaches (Blanchard et al., 1995; Weathers et al., 1999). An earlier study reports similarly high rates of sensitivity (.97; Foa et al., 1993)
- Agreement between the PSS-I and CAPS diagnoses of PTSD ranges 70–86 percent in clinical and nonclinical samples (Foa & Tolin, 2000)

- Convergent validity for the PSS-I among clinical and nonclinical samples is good, as evidenced by strong correlations with the CAPS and its domains (r scores range .63–.87; Foa & Tolin, 2000). Moreover, the correlations between the PSS-I and the SCID module for PTSD are equivalent to those between the CAPS and the SCID
- Among people who have severe mental disorders, subjective distress as indicated by the PSS-I is related to high risk behaviors, including drinking and attempted suicide (O'Hare et al., 2006)
- In support of the PSS-I's concurrent validity, among those with substance use and mental disorders, people diagnosed with PTSD using the PSS-I have significantly higher scores on the Addiction Severity Index for medical problems and higher rates of psychoticism, as measured by the Brief Symptom Inventory (Reynolds et al., 2005)
- The internal consistency of the PSS-I is quite good (alphas range .65–.86) in clinical and nonclinical samples (Foa & Tolin, 2000)
- The PSS-I has good interrater reliability across domains, with agreement ranging 94–99 percent (Foa et al., 2005; Foa & Tolin, 2000). An earlier study reported similar results (kappa = .91; Foa et al., 1993)

Concerns

- The PSS-I has not been studied extensively in criminal justice settings
- The generalizability of the PSS-I to a range of clinical settings has not yet been established
- Test-retest reliability of the PSS-I has not been widely examined

Availability and Cost

The PSS-I is a public domain instrument and can be downloaded without charge at the following site: http://www.istss.org/assessing-trauma/posttraumatic-symptom-scale-interview-version.aspx

Recommendations for Trauma/PTSD Screening, Assessment, and Diagnostic Instruments

Information regarding screening and diagnostic instruments for trauma and PTSD is based on a critical review of the literature and research comparing the efficacy of these instruments. Factors considered in recommending specific instruments include empirical evidence supporting the reliability and validity of the instrument, relative cost of the instrument, ease of administration, and use in the justice system. Although summaries of the instruments included research that was based on the DSM-IV criteria, recommendations are made considering the degree to which instruments align closely with the new DSM-5 criteria and allow for a more seamless transition to the new classification system. As noted before, although trauma/PTSD screening can be conducted by nonclinicians through use of standardized self-report instruments, screening staff should be knowledgeable about appropriate referral sources and the nature of trauma and PTSD. Offenders who screen positively as having significant problems related to trauma and PTSD should be referred for a comprehensive assessment to be conducted by a trained and licensed/certified mental health professional.

Based on the review of the literature and previously described considerations, the following screening instruments are recommended to examine the history of traumatic events and PTSD:

1. Either the Trauma History Screen (THS), or the Life Stressor-Checklist (LSC-R), or the Life Events Checklist-5 (LEC-5) to identify exposure to traumatic events.

(and)

2. The Posttraumatic Stress Disorder Checklist for DSM-5 (PCL-5) to identify trauma symptom severity.

This combined screen requires approximately 15–20 minutes to administer and score. For individuals who screen positive to the previous set of screens and for whom a more comprehensive assessment and/or diagnosis is needed, the following instruments are recommended:

1. The Posttraumatic Symptom Scale (PSS-I), which provides a current diagnosis of PTSD.

(or)

2. The Posttraumatic Diagnostic Scale (PDS), which serves as both a screen and diagnostic instrument.

(or)

3. The Clinician Assisted PTSD Scale for DSM-5 (CAPS-5).These assessment and diagnostic tools require approximately 25–30 minutes to administer and score.

Screening Instruments for Motivation and Readiness for Treatment

Several brief screening instruments have been developed to examine motivation and readiness for behavioral health treatment. These are sometimes used to identify individuals who are inappropriate for admission to substance use treatment, to flag issues that are important to address in early stages of treatment, and to monitor changes in motivation and readiness over the course of treatment. Although motivational screens are not always provided during the intake process, they may be used in different settings to determine readiness for change. Motivation and readiness for treatment have been found to predict treatment outcomes (Hiller, Knight, Leukefeld, & Simpson, 2002; Olver, Stockdale, & Wormith, 2011), including retention in and graduation from treatment programs, and may be particularly useful in matching individuals to different levels or "stages" of treatment. Motivation screens can be administered as a repeated measure to monitor progress over time.

A caveat to the use of motivational screens in matching people who have CODs to treatment in the criminal justice system is that this population is not typically motivated to participate in treatment and has a wide range of other psychosocial issues (e.g., housing, financial support) and personality factors (e.g., antisocial cognitions and attitudes) that may take precedence over treatment. Thus, motivation should not be viewed as a predicate for placing offenders in treatment. Instead, techniques aimed at increasing self-efficacy (setting small obtainable goals during treatment) and motivation (e.g., motivational interviewing techniques) for those who lack motivations and who are ambivalent about change can improve treatment outcomes in the justice system (CSAT, 2005b).

It is important to note several concerns regarding the validity of motivational screening instruments. First, not all of these instruments provide equivalent types of assessment of readiness for change, as some do not directly align with the stages of changes (e.g., SOCRATES), as defined by the transtheoretical model (TTM; Prochaska, DiClemente, & Norcross, 1992). Moreover, these instruments may provide variable results in assigning offenders to different "stages of change" or in identifying readiness for treatment, resulting in matching individuals to different levels of treatment. Thus, these measures should not be used as the primary tools to accomplish treatment matching.

Screening Instruments for Motivation and Readiness for Treatment

Circumstances, Motivation, Readiness, and Suitability Scale (CMRS)

The CMRS (DeLeon & Jainchill, 1986) was developed to assess risk for dropout from a therapeutic community (TC) program and to identify participants most likely to remain in substance use treatment. The CMRS is a 42-item scale that takes approximately 30 minutes to complete. The instrument has four subscales, Circumstances, Motivation, Readiness, and