Screening and Assessment of Co-Occurring Disorders in the Justice System Part 2

Quantum Units Education
Affordable. Dependable. Accredited.

www.quantumunitsed.com

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 7:

- 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 7. 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 7, 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 7 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 7 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 8:

- 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 8. 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 and scoring instructions can be downloaded at no cost at the following site, which includes information regarding scoring and interpretation: http://www.projectcork.org/clinical_tools/html/MAST.html

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

https://www.whitehouse.gov/sites/default/files/page/files/sbirt_fact_sheet_ondcp-samhsa_7-25-111.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 & Abojedi, 2010; Loutsiou-Ladd, Panayiotou, & 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 entire set of GAIN instruments (including the GAIN-SS) is available for $100, covering a period of 5 years of use. Multisite licenses are available. These arrangements include unlimited use of paper versions of the instrument. The GAIN-SS web version requires additional license agreements.

The GAIN instrument can be downloaded in both English and Spanish at the following website, but they are copyrighted. Unmarked paper versions of the instrument are part of the licensing package: http://www.gaincc.org/_data/files/Instruments percent20and percent20Reports/Watermarked percent20Instruments/GAINSS_3_0_ Watermarked.pdf

The GAIN-SS web version is available for $500 per year, which provides administration, scoring, and interpretative reports.

The GAIN-SS administration and instruction manual can be downloaded free of charge at the following site: https://www.assessments.com/assessments_documentation/gain_ss/GAIN-SS percent20Manual.pdf

This includes scoring and interpretation of the GAIN-SS paper instrument. Scoring instructions are located on the GAIN-SS instrument, and instructions for interpretation of GAIN-SS scores are located in the administration and instruction manual. 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.

Psychometric information across age groups can be found at the following site, including scales and variable descriptions for all versions of the GAIN: http://www.gaincc.org/psychometrics-publications/resources-for-evaluators-and-researchers/

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. The package can be purchased as paper instruments or as electronic computer-administered instruments. A licensing permission form for use of the MINI and MINI-Screen is provided. There is a one-time processing cost of $19.95. This cost is for individual use by students or private clinical practices. 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://www.medical-outcomes.com/index/mini

The Modified MINI-Screen can be downloaded at no cost at the following site, which includes instructions for scoring and interpretation: http://www.oasas.ny.gov/treatment/COD/documents/MMSTool.pdf

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 PDS starter kit costs approximately $60, which includes the PDS manual, test booklet, three answer sheets, and three administrations using Q software.

A hand-scoring starter kit costs approximately $67, which contains an administration manual, a test booklet, 10 answer sheets, 10 scoring worksheets, and a scoring instruction sheet.

Prices for scoring software vary according to the frequency of administration.

The PDS was discontinued; however, paper-based inventory will be sold until supplies run out. Information describing how to obtain the PDS can be found at the following site: http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAg510&Mode=summary

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 Suitability, that measure (1) external pressures to seek treatment; (2) internal reasons to seek change; (3) perceived need for treatment to achieve change; and (4) acceptance of the TC approach, reflected by the willingness to make major lifestyle changes, long-term commitment to an intensive treatment program, and rejection or exhaustion of other treatment modalities or options. A shortened 18-item version of the instrument (CMR) includes three subscales: Circumstances, Motivation, and Readiness.

Positive Features

- The CMRS is widely used among offenders (DeLeon, Melnick, Thomas, Kressel, & Wexler, 2000; Goethals, Vanderplasschen, Van de Velde, & Broekaert, 2012; Fiorentine, Nakashima, & Anglin, 1999; Melnick, DeLeon, Thomas, Kressel, & Wexler, 2001) and people with substance use disorders (Battjes, Gordon, O'Grady, Kinlock, & Carswell, 2003; DeLeon, Melnick, & Cleland, 2010; Gholab & Magor-Blatch, 2013; Najavits et al., 1997)
- The CMRS consistently predicts retention and entry into prison-based TCs and entry into aftercare TCs following release from custody (DeLeon, Melnick, Thomas, Kressel, & Wexler, 2000)
- The abbreviated CMR instrument predicts involvement in substance use aftercare treatment following release from prison (Melnick, DeLeon, Hawke, Jainchill, & Kressel, 1997)
- Among participants in the Drug Abuse Treatment Outcome Study (DATOS), scores on the treatment readiness scale of the CMRS predict treatment retention across treatment settings, supporting the predictive validity of the measure (Joe, Simpson, & Broome, 1999)
- The CMR is positively related to aftercare involvement in prisoners enrolled in TCs, and higher scores on the CMR predict aftercare entry and lower reincarceration rates at a 1-year follow-up (Melnick et al., 2001)
- Among offenders enrolled in TC programs, treatment motivation scores on the CMR predict treatment readiness (Morgen & Kressel, 2010)
- Among offenders in TC programs, treatment motivation as indexed by the CMRS is related to environmental factors,

such as understanding the rules of conduct and treatment goals (Goethals et al., 2012)

- Treatment motivation as indexed by the CMR is directly related to treatment alliance, treatment participation, and treatment outcomes (Melnick et al., 2001)
- The CMRS is useful in predicting 30-day retention in long-term TC treatment in the community (DeLeon et al., 1994)
- Young (2002) found that external factors measured by the Circumstances scale of the CMRS predicted 90-day retention of criminal justice clients in community-based residential treatment programs, while the Readiness scale of the CMRS predicted 180-day retention
- Melnick et al. (1997) found that age was significantly correlated with scores on the CMRS and that the instrument successfully predicted short-term retention rates in TC treatment across age groups
- DeLeon, Melnick, Kressel, and Jainchill (1994) found that CMRS scales are more effective predictors of 30-day and 10-month treatment retention than a range of demographic and background variables, including legal status
- People mismatched to treatment in the DATOS had significantly lower CMR treatment motivation scores at baseline in comparison to those who were properly matched to treatment (DeLeon et al., 2010)
- Higher motivation for mental health treatment as indexed by the CMR predicts greater adherence to treatment among psychiatric patients (Magura, Mateu, Rosenblum, Matusow, & Fong, 2014)
- The CMR has good predictive utility for treatment outcomes across race and ethnicity (DeLeon, Melnick, Schoket, & Jainchill, 1993)
- Reliability of the CMRS total score as measured by Cronbach's alpha is .84 (Melnick et al., 2001), and reliabilities for individual scale scores range from .53 for the Circumstances scale to .84 for the Readiness scale
- The CMRS has good internal consistency (alphas = .84–.87; .67–.83; DeLeon et al., 1994; Goethals et al., 2012, Melnick, 1999)

Concerns

- CMRS scores vary significantly for offenders of differing intellectual functioning (Van de Velde, Broekaert, Schuyten, & Van Hove, 2005)
- The CMRS items are related to TCs, and thus, the instrument may not generalize to other treatment settings for assessing circumstances, motivation, and readiness for change (Groshkova, 2010; Zemore & Ajzen, 2014)
- The validity of the CMRS has not been examined among individuals with CODs
- The CMRS has not been thoroughly evaluated to determine its usefulness in predicting retention in jail or community-based offender treatment programs
- Circumstances scale scores have low reliability (Van de Velde et al., 2005)
- The Circumstances scale may consist of two factors, Pressures to Enter Treatment, and Pressures to Leave Treatment (DeLeon et al., 2000), thus explaining difficulties related to low reliability. Caution should be used when interpreting this scale

Availability and Cost

The CMRS manual and instruments can be obtained free of charge at the following site: http://www.emcdda.europa.eu/html.cfm/index3597EN.html

Readiness for Change Questionnaire (RCQ)

The RCQ (Rollnick, Heather, Gold, & Hall, 1992) is a 12-item measure based on the transtheoretical "stages-of-change" model, developed by Prochaska and DiClemente (1992). The instrument was originally developed to identify specific stages of change among heavy drinkers who are not seeking treatment, but it has been used far more broadly among a range

of substance-involved populations. The RCQ-CV (clinician's version) consists of three scales, Pre-contemplation, Contemplation, and Action, each consisting of four items. Item responses are provided on a five-point scale ranging from "strongly agree" to "strongly disagree," with higher scores on the RCQ representing greater willingness to change. The 15-item RTCQ-TV (treatment version) was designed for individuals in treatment or who are seeking treatment (Share, McCrady, & Epstein, 2004) and is used to determine the level of readiness to engage in treatment and to assist in treatment planning. A revised 12-item version of the RTCQ-TV is also available (Heather & Honekopp, 2008). Both the RCQ-CV and RTCQ-TV take approximately 2–3 minutes to administer, are designed for both adolescents and adults, and are available in the public domain. The RCQ has been adapted to measure readiness to change in other areas, such as violent behavior, criminal behaviors, and anger problems. Neither instrument requires training to administer or score.

Positive Features

- The RCQ is brief to administer
- The self-administered format of the RCQ is advantageous for use in hospital and other settings in which there is limited time to compile information (Rollnick et al., 1992). The RCQ has been used with several offender populations (Casey, Day, Howells, & Ward, 2007; Day et al., 2009; McMurran et al., 1998; Watt, Shepherd, & Newcombe, 2008) and with people with substance use disorders (Freeman et al., 2005; Heather, Luce, Peck, Dunbar, & James, 1999; Gregoire, & Burke, 2004; Share, McCrady, & Epstein, 2004; Wells-Parker, Kenne, Spratke, & Williams, 2000)
- The RCQ has been adapted for use with offenders (Readiness to Change Offending, RCOQ) to address motivation to change criminal behaviors (McMurran et al., 1998)
- The RCQ is related to a newly developed offender instrument that examines readiness for change, the Corrections Victoria Treatment Readiness Questionnaire (CVTR), and demonstrates moderate to strong correlations with the CVTR scales (Casey et al., 2007)
- The RCQ has been adapted to measure readiness to change violent behaviors among offenders and is correlated with another treatment readiness scale, the Violence Treatment Readiness Questionnaire (VTRQ; Day et al., 2009)
- Convergent validity of the RCQ among people involved in substance use treatment is supported by correlations with another well-validated measure of readiness for change, the URICA (r scores range .39–.56; Heather et al., 1999)
- Violent offenders who received no intervention were more likely to be in the pre-contemplation stage for changing drinking behaviors compared to those receiving a treatment intervention, supporting the validity of the RCQ in assessing readiness for change (Watt et al., 2008)
- Convergent validity of the instrument is also indicated among people with substance use disorders, in which RCQ scores indicating pre-contemplation, contemplation, and action stages are related to scores from the URICA, another well-validated measure of readiness for change (Napper et al., 2008)
- Support for the concurrent validity of the RCQ is provided among a substance-involved sample, in which people scoring in the pre-contemplation range showed significantly more injection drug use relative to those in the action stage. People scoring in the pre-contemplation range also remained in treatment for fewer weeks than those scoring in the contemplation range (Napper et al., 2008)
- People who had received substance use treatment were more likely to receive RCQ scores in the action stage. Moreover, those who had better treatment outcomes were more likely to be in the action or

contemplation stage compared with those who had poor treatment outcomes, supporting the validity of the measure for assessing readiness for change (Heather et al., 1999)

- The RCQ's validity is supported among a sample of offenders who were court-mandated to outpatient substance use treatment because they were more likely to be in the action or contemplation stage compared to those not receiving treatment, even after controlling for level of substance use problems (Gregoire & Burke, 2004)
- In a sample of repeat DUI offenders, those determined to be in the contemplation stage by the RCQ for changing level of alcohol consumption had higher self-efficacy for controlling their drinking and had lower levels of alcohol consumption relative to those in the pre-contemplation stage (Freeman et al., 2005). Another study (Wells-Parker et al., 2000) indicates that those determined to be in the action stage by the RCQ for reducing drinking and driving behaviors have lower rates of criminal recidivism. These studies support the concurrent validity of the RCQ instrument
- Several other studies demonstrate the discriminant and convergent validity of the RCQ in measuring readiness for change among DUI offenders (Freeman et al., 2005; Wells-Parker & Williams, 2002)
- The RCQ has good predictive validity for changes in drinking behavior over time (Share, McCrady, & Epstein, 2004)
- The revised RCQ-TV shows a good fit with a three-factor structure, supporting the three scales of the RCQ-TV (Heather & Honekopp, 2008)
- The revised RCQ-TV total scale score shows good internal consistency (alpha > .70), particularly for the Action scale (alpha = .85; Heather & Honekopp, 2008). Previous studies indicated that the RCQ has satisfactory internal consistency, with Cronbach's alphas of .73 for the Pre-contemplation subscale, .80 for the Contemplation scale, .85 for the Action scale (Rollnick et al., 1992; Napper et al., 2008), and .71 for the entire scale (Day et al., 2009)
- Test-retest reliability for the RCQ scales has been found to be satisfactory (Rollnick et al., 1992), with correlations of .82 (Pre-contemplation), .86 (Contemplation), and .78 (Action). Test-retest reliability of the RCQ among those enrolled in substance use treatment is quite good over a 3-day interval (r scores range .69–.86 across RCQ scales; Heather et al., 1999). Good test-retest reliability of the revised RCQ-TV has also been demonstrated among people enrolled in alcohol treatment (r scores range .76–.88) for all stages of change, over a 3-month interval (Heather & Honekopp, 2008)

Concerns

- The validity of the RCQ has not been widely studied among offenders and additional research on its psychometric properties among this population is needed
- Little evidence has been found to support concordance between interviewer-determined stage of change and stage of change assessed by the RCQ (kappas range .08–.45; Addington, El-Guebaly, Duchak, & Hodgins, 1999)
- The internal consistency of the RCQ may be somewhat low (alpha = .69; Casey et al., 2007), particularly for the Pre-contemplation scale (alpha = .68; Napper et al., 2008) and the Contemplation scale (alpha = .60–65; Heather et al., 1999; Napper et al., 2008)
- The revised RCQ-TV shows low internal consistency for the Pre-contemplation (alpha = .66) and Contemplation scales (alpha = .66; Heather & Honekopp, 2008)
- The RCQ (McMurran et al., 1998) shows low internal consistency for the Pre-contemplation (alpha = .60) and Contemplation (alpha = .49) scales

Availability and Cost

The RCQ is copyrighted but is available free of charge.

The RCQ–CV measures and related materials can be accessed at no cost at the following site, which includes information regarding scoring, interpretation, and reliability and validity of the instrument: http://www.addiction.ucalgary.ca/researchers/instruments

The revised RCQ-TV can be obtained at the following site, as part of a manuscript describing the validity of the instrument. Scoring and interpretation guidelines are provided in the manuscript appendices: http://www.researchgate.net/publication/232067129_A_revised_edition_of_the_Readiness_to_Change_Questionnaire_Treatment_Version

Stages of Change Readiness and Treatment Eagerness Scale (SOCRATES)

The SOCRATES provides a family of instruments designed to examine readiness for change among substance-involved populations, according to the "stages-of-change" model (Prochaska & DiClemente, 1992). The SOCRATES was developed through funding by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) and is a "public domain" instrument. The original instrument provided five separate scales corresponding with the stages-of-change model, while a more recent factor analysis of the SOCRATES has led to the development of three scales: Ambivalence, Recognition, and Taking Steps, each of which reflects different stages of motivation and readiness for treatment. The SOCRATES is often used as a repeated measure to assess change in motivation over time related to involvement in motivational interviewing interventions and substance use treatment. The 19-item version has the following recommended cut-off scores for the Recognition scale: low scores are ≤30, medium scores are 31–34, and high scores are ≥35. For the Ambivalence scale, cut-offs for low scores are ≤ 13, medium scores are 14–16, and high scores are ≥ 17.

Several versions of the SOCRATES have been developed for different populations, including the following:

- 8D/A (19 items)—drug and alcohol questionnaire for clients
- 7A-SO-M (32 items)—alcohol questionnaire for significant others of males
- 7A-SO-F (32 items)—alcohol questionnaire for significant others of females
- 7D-SO-F (32 items)—drug and alcohol questionnaire for significant others of females
- 7D-SO-M (32 items)—drug and alcohol questionnaire for significant others of males

Positive Features

- The instrument is brief to administer and is easily scored
- The SOCRATES has been used with a range of offender populations (Brocato & Wagner, 2008; Evans, Huang, & Hser, 2011; Morris & Moore, 2009; Prendergast et al., 2009; Vanderburg, 2003) and people with substance use disorders (Gossop, Stewart, & Marsden, 2007; Kelly, Finney, & Moos, 2005; Napper et al., 2008; Zhang, Harmon, Werkner, & McCormick, 2004) and is commonly used with offenders to assess readiness for change (Gunter, Antoniak, 2010)
- The Recognition and Taking Steps scales of the SOCRATES have been identified as important factors in motivation for change and are reliably distinguishable in the beginning of treatment (Carey, Maisto, Carey, & Purnine, 2001; Isenhart, 1997; Miller & Tonigan, 1996)
- Scores on the SOCRATES are correlated with attempts to quit both alcohol and drug use (Henderson, Saules, & Galen, 2004; Isenhart, 1997; Zhang et al., 2004)

- In support of the concurrent validity of the SOCRATES 19-item version, people scoring higher on the Recognition scale have greater drug use and symptoms of depression and anxiety than people scoring higher on the Taking Steps scale (Gossop et al., 2007)
- Also supporting the concurrent validity of the SOCRATES 19-item instrument, people with substance use disorders who spent a shorter amount of time in drug treatment were more likely to score at the Pre-contemplation stage compared to those scoring at the Determination and Action stage. Those scoring at the Action stage also had significantly fewer days of drug use than people who were at the Pre-contemplation and Determination stage (Napper et al., 2008)
- In a sample of nonviolent offenders who had committed drug crimes, the SOCRATES Recognition scale predicted arrests within the past 12 months, and both the Ambivalence and Taking Steps scales predicted drug arrests during the past 12 months (Prendergast et al., 2009)
- Among offenders with alcohol use problems, those who received a motivational interviewing intervention scored higher on the Recognition scale of the SOCRATES, in addition to change from the Pre-contemplation to Contemplation stage of change, as measured by the University of Rhode Island Change Assessment Scale (URICA) and RCQ, supporting the convergent validity of the SOCRATES 19-item instrument (Mann, Ginsburg, & Weekes, 2002)
- In a study of offenders who were court-mandated to substance use treatment, those who remained longer in treatment had significantly higher total scores on the SOCRATES compared to dropouts, supporting the validity of the measure (Brocato & Wagner, 2008). The SOCRATES total score also predicted length of treatment stay, and the Recognition scale predicted therapeutic alliance and length of treatment stay across groups differing by race/ethnicity and type of primary drug use
- The SOCRATES ambivalence scale shows reliable and clinically significant change from pre to post-treatment among offenders, supporting its ability to assess change in motivation over time (Morris & Moore, 2009)
- In a sample of substance-involved military personnel, the SOCRATES Ambivalence, Recognition, and Taking Steps scales are related to commitment to abstinence, disease attribution, and powerlessness, as measured by the Addiction Treatment Attitude Questionnaire (ATAQ; Mitchell & Angelone, 2006). The same study found that the SOCRATES Ambivalence scale is related to treatment completion, supporting the concurrent validity of the measure
- Internal consistency coefficients for the SOCRATES are quite good, with alphas ranging .81.–93 for the Recognition scale, .84–88 for Taking Steps, and .71 for Ambivalence (Gossop et al., 2007; Mitchell, Francis, & Tafrate, 2005; Brocato & Wagner, 2008)
- The test-retest reliability of the SOCRATES is quite high among correctional populations (Peters & Greenbaum, 1996). Test-retest reliability (Miller & Tonigan, 1996) of the SOCRATES over a 2-day interval is also quite good across different scales, including Ambivalence (r score = .83), Recognition (r score = .99), and Taking Steps (r score = .93)
- The SOCRATES Recognition scale has moderately good sensitivity and specificity in identifying substance-dependent offenders (Peters & Greenbaum, 1996)

Concerns

- The validity of the SOCRATES has not been widely examined among individuals with CODs
- The SOCRATES may contain some confusing and ambiguous language, which

can detract from effective assignment of individuals to different stages of change. The determination of stages of change by the SOCRATES is not always consistent with stages of change determined by other measures, such as by the RCQ (Burrowes & Needs, 2009; Lechner, Brug, De Vries, van Assesma, & Muddle, 1998; Littell & Girvin, 2002; Williamson, Day, Howells, Bubner, & Jauncey, 2003)

- The SOCRATES may not be able to clearly distinguish among the five stages of change (DiClemente, Schlundt, & Gemmell, 2004)
- Although a study conducted by Nochajski and Stasiewicz (2005) did not support the use of the SOCRATES with DUI offenders, the Ambivalence and Recognition subscales were found to be associated with binge drinking
- The SOCRATES 19-item version may not detect changes in motivation among drug-involved offenders who received a motivational interviewing intervention, as well as the RCQ (Vanderburg, 2003)
- Not all subscales of the SOCRATES may be useful in predicting treatment retention. For example, the Ambivalence and Taking Steps scales were not found to predict length of stay in treatment among offenders (Brocato & Wagner, 2008)
- The SOCRATES may be more useful when used in combination with the URICA to assess readiness to change (DiClemente et al., 2004)
- In a review of the existing literature, DiClemente, Schlundt, and Gemmell (2004) found only modest support for the predictive validity of the SOCRATES
- Research provides support for both two- and three-factor structures for the SOCRATES (Demmel, Beck, Richter, & Reker 2004; Figlie, Dunn, & Laranjeira, 2005; Mitchell et al., 2005) and indicates that the number of items could be reduced
- The internal consistency of the SOCRATES is low when used to determine readiness for change via stages of change (Hodgins, 2001) that include Pre-contemplation, Contemplation, Determination, and Maintenance, with alphas < .61 (Napper et al., 2008)
- Internal consistency of the Ambivalence scale is low (alpha = .38; Gossop et al., 2007)
- The SOCRATES exhibits low agreement with other validated measures of readiness to change, such as the URICA and RCQ, across the various stages of change (<40 percent agreement; Napper et al., 2008)

Availability and Cost

The SOCRATES is available free of charge at the following site: http://casaa.unm.edu/inst/socratesv8.pdf

Texas Christian University Motivation Form (TCU MOTForm)

The TCU MOTForm is a 36-item instrument that examines not only readiness for change but also motivation and readiness for treatment. Items are worded specifically for drug-involved populations. The instrument includes five scales, including Problem Recognition (PR), Desire for Help (DH), Treatment Readiness (TR), Pressures for Treatment (PT), Treatment Needs (TN), and Accuracy (Attentiveness). Accuracy is a single item that identifies whether the respondent is paying attention while completing the measure. Respondents indicate how strongly they agree or disagree with the statement on a one (disagree strongly) to five (agree strongly) scale. Higher scores indicate higher levels of motivation for treatment. The TCU MOTForm can be used prior to treatment to examine motivation and readiness for change and as a repeated measure to monitor change over time. It was developed for criminal justice settings.

Positive Features

- The TCU MOTForm is brief to administer, score, and interpret

- The TCU MOTForm was developed for use in criminal justice settings
- A greater desire for help (DH) as measured by the TCU MOTForm is related to greater treatment participation (Joe, Simpson, Greener, & Rowan-Szal, 1999)
- Treatment readiness (TR) as measured by the TCU MOTForm is related to improved post-treatment outcomes (Joe, Simpson, Greener et al., 1999; Simpson, Joe, Greener, & Rowan-Szal, 2000)
- Among offender and community-based treatment samples, the TCU MOTForm scales of PR, DH, and TR are correlated with treatment engagement, satisfaction, counselor rapport, and peer support (Joe, Simpson, & Broom, 1999; Pankow et al., 2012; Simpson et al., 2000; Simpson et al., 2012). The DH, TR, and TN scales also predict significant variance in treatment participation, supporting the predictive validity of the scales (Simpson et al., 2012)
- Across gender groups among offender samples, people with higher scores on the TCU MOTForm have higher levels of treatment participation, supporting the validity of the measure (Simpson et al., 2012)
- Across prison and community-based treatment settings, the TCU MOTForm scales are related to scales from the Addiction Severity Index (ASI). Specifically, the PR, DH, and TN scales are positively related to higher scores on the psychiatric, medical, legal, drug, alcohol, and employment scales of the ASI, supporting the concurrent validity of the TCU MOTForm (Pankow et al., 2012)
- Among offenders, higher scores on the TCU MOTForm (particularly the DH, TR, and TN scales) are negatively correlated with criminal thinking scales such as power orientation, coldheartedness, criminal rationalization, and entitlement (Garner, Knight, Flynn, Morey, & Simpson, 2007), supporting the concurrent validity of the TCU MOTForm
- An exploratory factor analysis of the MOTForm instrument shows a good fit for each scale, as evidenced by a single factor structure for each subscale (Simpson et al., 2012)
- The TCU MOTForm has good internal consistency for each scale, PR (alpha = .87–.90), DH (alpha = .66–.81), TR (alpha = .75–.84), and TN (alpha = .64), in both community and criminal justice settings (Garner et al., 2007; Simpson et al., 2012; Simpson & Joe, 1993)
- The test-retest reliability of the TCU MOTForm is quite high over a 2-week interval (r scores range .74–.88)

Concerns

- Additional research is needed regarding the predictive validity of the TCU MOTForm in criminal justice and community settings and with populations who have CODs
- The TCU MOTForm scales of TN and DH may have lower internal consistency (alpha = .64–.67) in comparison to the other scales (Garner et al., 2007; Simpson et al., 2012)
- A confirmatory factor analysis provides inconsistent results to support a single factor structure for each scale, and some scales may be multidimensional in nature. The authors of the MOTForm report that these results may be due to combining results obtained prior to treatment with those obtained during the course of treatment, at which time the meaning of motivation and readiness may have changed with treatment progress (Garner et al., 2007; Simpson et al., 2012)

Availability & Cost

The TCU MOTForm is available in the public domain, and the instrument along with materials related to scoring and interpretation can be found at the following site: http://ibr.tcu.edu/forms/treatment-motivation-scales/

University of Rhode Island Change Assessment Scale (URICA)

The URICA (DiClemente & Hughes, 1990; McConnaughy, Prochaska, & Velicer, 1983) includes 24-, 28-, and 32-item versions of the self-report questionnaire examining motivation and readiness for treatment. The 32-item URICA consists of four scales made up of 8 items each, while the 28-item and the 24-item versions have four scales consisting of 7 and 6 items, respectively. The 24-item version has been adapted to those with CODs (URICA-M). The URICA-M uses simpler language, defines problems identified by the instrument with the respondent, and can be administered as an interview for those who have problems related to literacy or sight. A 12-item version of the URICA is available that examines readiness to change drinking behaviors and includes four scales. The four scales were developed to examine each of the theoretical stages of change (Pre-contemplation, Contemplation, Action, and Maintenance) related to individual motivation for treatment (DiClemente & Prochaska 1982, 1985; Prochaska & DiClemente, 1992).

The URICA appears to identify two distinctive subtypes: pre-contemplation and contemplation/action (Blanchard, Morgenstern, Morgan, Labouvie, & Bux, 2003; Edens & Willoughby, 1999, 2000). Readiness to change (RTC) can be calculated from the URICA instrument by subtracting mean Pre-contemplation scores from Contemplation, Action, and Maintenance scores (Connors et al., 2000; Project MATCH Research Group, 1997). A Contemplative Action score (CA) can be calculated by subtracting mean Contemplation scores from Action scores (Pantalon, Nich, Frankforter, & Carroll, 2002). The following cut-off scores may be appropriate for the general population: < 8 to be classified as "Pre-contemplators," 8–11 to be classified as "Contemplators," and 11–14 to be classified as "Preparators into Action Takers." URICA scale scores may vary across different settings and stages of change in the particular settings. Thus, use of the URICA to classify individuals to various stages of change should consider profiles generated from the particular setting that correspond with stages of change in that setting. The URICA differs from the SOCRATES and several other motivational screens in that it does not directly ask about motivation for alcohol or drug treatment but instead presents questions in a more general manner. The URICA does not require clinical training to administer or score.

Positive Features

- The URICA is brief to administer and score
- The URICA has been used with offender populations (Alexander & Morris, 2008; Brodeur, Rondeau, Brochu, Lindsay, & Phelps 2008; Levesque, Gelles, & Velicer, 2000; Polaschek, Anstiss, & Wilson, 2010; Tierney & McCabe, 2004), people with substance use disorders (Callaghan et al., 2008; Budney, Higgins, Radnovich, & Novy, 2000; Budney, Moore, Rocha, & Higgins, 2006; Field, Adinoff, Harris, Ball, & Carroll, 2009; Jungerman, Andreoni, & Laranjeira, 2007), and those with CODs (Bellack et al., 2006; Kinnaman, Bellack, Brown, & Yang, 2007; Nidecker, DiClemente, Bennett, & Bellack, 2008)
- The URICA has been adapted for domestic violence offenders (URICA-DV), and the instrument properties are consistent with the original URICA four-scale model. The URICA-DV shows good psychometric properties and is correlated with domestic violence behaviors such as history of violence, blame, and changing violent behaviors (Levesque et al., 2000)
- The URICA-DV demonstrates good concurrent validity (Alexander & Morris, 2008) such that those determined to be in later stages of change (higher scores on contemplation, action and maintenance) report less psychological aggression against their partner during the previous 6 months
- The URICA's validity in assessing readiness for change is demonstrated in outpatient substance use treatment

settings (Field, Duncan, Washington, & Adinoff, 2007), where RTC scores are correlated with increased anger problems and experience of recent life stressors, suggesting that RTC reflects the desire to change and seek help. In these settings, CA scores are negatively correlated with alcohol problems and anxiety, indicating that CA may reflect commitment to change substance use behaviors. Three studies involving outpatient substance use treatment participants (Budney et al., 2000; Budney et al., 2006; Jungerman et al., 2007) found that URICA scores were negatively correlated with marijuana use and related problems after treatment, supporting the concurrent validity of the URICA (Callaghan et al., 2008)

- Support for the convergent and concurrent validity of the URICA has been shown in outpatient treatment settings, in which higher RTC scores are correlated with more severe drug and alcohol problems (Field et al., 2009), while higher CA scores are associated with less severe alcohol and drug use problems and less severe familial and medical problems (Field et al., 2009)
- The validity of the URICA has also been demonstrated among people with CODs. Among this population, higher psychiatric distress is correlated with endorsement of negative aspects of drinking and higher scores on the Maintenance scale of the URICA, indicating greater difficulties in attempts to maintain sobriety (Velasquez, Carbonari, & DiClemente, 1999)
- In support of the convergent validity of the URICA among people who have CODs, the URICA-M is correlated with other measures of change, such as the Process of Change Scale (POC; DiClemente, Carbonari, Addy, & Velazquez, 1996) and its subscales and the "cons" of drug use from the Decisional Balance Scale (DBS; Velicer, DiClemente, Prochaska, & Brandenburg, 1985). The relationship between the POC and the URICA-M are strongest among depressed individuals (Nidecker et al., 2008)
- The URICA is able to discriminate between readiness to change among people who are alcohol dependent, with and without co-occurring depression (Shields & Hufford, 2005)
- The concurrent and convergent validity of the URICA in predicting change in criminal behaviors among offenders is supported by high correlations (r score = .80) with the Criminogenic Needs Inventory (CNI; Coebergh, Bakker, Anstiss, Maynard, & Percy, 2001) and low correlations (r score = -.42) with an inventory of deceptive behaviors, the Balanced Inventory of Desirable Responding (BIDR; Paulhus, 1998; Polaschek et al., 2010)
- The URICA has good psychometric properties in predicting change in criminal behaviors (Field et al.,2009; Tierney & McCabe, 2004; Polaschek et al., 2010)
- The URICA-M demonstrates good psychometric properties as a unitary scale among those with CODs (Nidecker et al., 2008), as the Pre-contemplation scale is negatively correlated with other scales (-.25 to -.30), while Contemplation, Action, and Maintenance scales are positively correlated with each other (r scores range .48–.80)
- The URICA has good internal consistency among people with CODs (Pantalon & Swanson, 2003). When applied to changing criminal behavior among offenders, internal consistency is acceptable for the 32-item URICA (alpha = .82) and across scales of Pre-contemplation (alpha = .75–.83), Contemplation (alpha = .60–90), Action (alpha =.81–.93), and Maintenance (alpha =.89–.90; Polaschek et al., 2010; Tierney & McCabe, 2004). Internal consistency of the URICA is also good when applied to changing substance use behaviors, for scales of Pre-contemplation (alphas range .73–.80), Contemplation (alphas range .72–.90), Action (alphas range .71–.81), and Maintenance (alphas

range .67–.74; Field et al., 2009; Nidecker et al., 2008)

- The URICA has good reliability, with estimates ranging .79–.88 (Carey, Purine, Maisto, & Carey, 1999). Reliability estimates for the URICA are .68–.85 among alcohol, opiate, cocaine, and nicotine-dependent individuals (Blanchard et al., 2003)

Concerns

- Additional research is needed to establish the validity of the URICA with offenders
- Among people with CODs, the URICA may not predict levels of treatment participation, treatment retention, dropout, or other treatment outcomes (Bellack et al., 2006; Kinnaman et al., 2007)
- Research examining the validity of the URICA has yielded mixed results. Studies involving people with alcohol user disorders and psychotherapy clients provide support for the validity of the URICA's four scales, but studies involving people with other drug use disorders do not provide similarly strong support (Carey et al., 1999; DiClemente et al., 2004)
- Although good concurrent validity was found for the four URICA scales and for the overall score, one study found that neither the scales, nor the overall score successfully predicted treatment outcome (Blanchard et al., 2003)
- The URICA produces scores related to four stages of change. However, these aren't precisely aligned with the most recent transtheoretical model of change (Prochaska et al., 1992), in which the Preparation stage has been eliminated due to poor fit with the instrument's underlying factor structure (Polaschek et al., 2010)
- When applied to changing criminal behavior, the four-factor structure of the URICA may be more accurately represented by deletion of items 2, 8, and 24, based on findings of improved internal consistency and fit across the various scales (Polaschek et al., 2010). The internal consistency of the Contemplation scale may also be low among offenders when applied to changing criminal behaviors (alpha = .90; Polaschek et al., 2010)

Availability and Cost

The URICA is available free of charge. The URICA instruments and materials describing scoring and interpretative guidelines can be found at the following site: http://habitslab.umbc.edu/urica/

Recommendations for Motivational Screening Instruments

Information regarding motivational screening instruments is based on a critical evaluation of the literature, including comparative research examining the efficacy of these instruments. Important factors in determining the utility of motivational screens include empirical evidence supporting the reliability and validity of the instruments, cost of the instruments, and ease of administration and scoring within the criminal justice settings. Motivation can also be focused on a variety of domains (e.g., substance use, mental health, criminal justice involvement). Specific to the area of motivational screening, instruments recommended are those that closely align with the transtheoretical model (TTM) and stages of change and that have demonstrated validity within the criminal justice system. The following instruments are recommended:

1. The Texas Christian University Motivation Form (TCU-MOTForm). This instrument is unique in identifying not only readiness to change but also variables related to motivation and treatment engagement, including problem recognition, desire for help and treatment readiness.

(or)

2. The University of Rhode Island Change Assessment Scale (URICA), which provides efficient identification of readiness to change

and directly maps onto four out of the five transtheoretical stages of change. The URICA-M is specifically designed for people with CODs and provides simpler language and a shorter administration time.

Both of these instruments have been examined in the criminal justice system and/or among people with CODs. The URICA is recommended for settings in which it is important to determine readiness to change, while the TCU-MOTForm can also be used to assess issues related to treatment engagement. Each of these measures requires approximately 10–15 minutes to administer and score.

Assessment Instruments for Substance Use and Treatment Matching Approaches

The use of assessment to match justice-involved individuals to appropriate levels of behavioral health services has been recognized as among the most fundamental of evidence-based approaches (CSAT, 2005b). The goal of treatment matching is to provide an individualized examination of a range of mental and substance use disorders and other related psychosocial problems to assist in matching offenders to appropriate levels of services. Triage to appropriate services is particularly important among offenders who have CODs, as mental and/or substance use disorders often go undetected, and this population is often mismatched to less intensive services than are needed. This section describes several treatment matching approaches, as well as specific assessment instruments to assist in matching offenders with CODs to specific services. Matching approaches include the Risk-Need-Responsivity model and the American Society of Addiction Medicine's Patient Placement Criteria (ASAM PPC). Both of these approaches provide detailed frameworks for assessing substance use disorders, mental disorders, and other areas related to placement in treatment and supervision services. Assessment instruments and treatment matching approaches should be administered by mental health professionals with advanced clinical training related to mental and substance use disorders, diagnosis, referral to treatment, and treatment planning. Several of the structured and standardized self-report assessment instruments described in this section can be administered by nonclinicians, although staff should be knowledgeable about appropriate referral sources.

Specific assessment instruments described in this section include the Addiction Severity Index (ASI), the Timeline Followback (TLFB), and the TCU Correctional Justice instruments (TCU CJ).

Identifying Gaps in Offender Services

Despite the availability of several treatment matching approaches and instruments, there are significant challenges in matching offenders who have CODs to appropriate levels of care, due to the lack of available treatment and supervision services in many jurisdictions. Belenko & Peugh (2005) developed a protocol to identify gaps in treatment services (primarily substance misuse services) within correctional systems. In order to identify offenders' treatment needs, guidelines were developed to assess substance use severity, recency of substance use problems, consequences of substance use, and other psychosocial and health problems. The second step involved surveying available correctional treatment resources and categorizing them according to the following schema: (1) no treatment (low level of drug use, no drug related consequences), (2) short-term intervention (self-help, motivational interviewing), (3) outpatient treatment (individual or group counseling), and (4) residential treatment (separate housing, long-term intensive treatment for those with several drug related consequences and frequent drug use). Using this protocol, they compared offenders' treatment needs with actual treatment received within a large correctional sample. Results indicated that approximately a third of male and female prisoners needed residential treatment, and approximately 16–18 percent needed outpatient treatment. A survey

of correctional institutions revealed that only 19 percent of males and 23 percent of females actually received substance use treatment, and of those receiving treatment, about a third received only drug education or self-help groups (e.g., AA/NA). These findings highlight the importance of using a formal assessment approach to identify needs of offenders and to provide matching to specific levels of treatment services, and challenges in treatment matching within an environment that often includes scarce treatment resources and with a population that has pronounced treatment needs (e.g., offenders with CODs).

Treatment Matching Approaches

Risk-Need-Responsivity Model

The Risk-Need-Responsivity (RNR) model identifies the importance of identifying "criminogenic needs" of offenders that are related to recidivism and using this information to match offenders to different levels of treatment and supervision (Andrews & Bonta, 2010b). The "risk principle" encourages assessment of criminal risk to ensure that intensive resources (e.g., CODs treatment, substance use treatment) are reserved for offenders who are at moderate to high risk levels. Key predictors of criminal risk include "static" or unchanging factors (e.g., age, age at first arrest, number of prior arrests/convictions) and "dynamic" or changeable factors, such as criminal attitudes and beliefs, criminal peers, substance use problems, employment, education, family problems, and lack of prosocial leisure skills.

The most important predictors of criminal risk are past criminal behavior and antisocial attitudes, beliefs, and peers, although substance use problems also represents an important risk factor. Although mental illness is not an independent risk factor for recidivism, offenders who have mental disorders are at elevated criminal risk due to having high levels of criminogenic needs (e.g., ingrained criminal belief systems, poor employment history, lack of education). Offenders who have CODs are at particularly high risk for recidivism and should be a priority population for programming and specialized supervision (Drake, 2011). A range of risk assessment instruments has been developed that examines both static and dynamic risk factors and provides overall criminal risk scores and recommendations for placement in different levels of treatment and supervision. Various risk assessment instruments are described in the "Risk Assessment" section of this monograph.

The RNR model asserts that dynamic risk factors ("criminogenic needs") should be targeted in individualized assessment and offender programming to most effectively reduce recidivism. Many offender programs, including those providing treatment for CODs, do not address a range of these criminogenic needs, and as a result, are less likely to reduce recidivism (Lowenkamp & Latessa, 2005). Research indicates that there is a cumulative effect in addressing criminogenic needs, resulting in a linear relationship between the number of needs addressed in offender treatment and supervision and positive outcomes related to recidivism (Bonta & Andrews, 2010; Carey & Waller, 2011).

The RNR model also indicates the need to address "responsivity" in offender programs, referring to factors that influence the offender's engagement in evidence-based treatment (e.g., services that address dynamic risk factors/criminogenic needs). Responsivity factors include mental health problems, need for gender-specific services, history of trauma/PTSD, need for culturally sensitive programming, and various disabilities. If unaddressed, responsivity factors can undermine engagement, retention, and outcomes in offender treatment and supervision.

Consideration of the three components of the RNR model (risk, criminogenic needs, responsivity) provides a very useful framework for matching offenders to different types and intensity of treatment and supervision. Appropriate matching based on these principles leads to reductions in

recidivism and other positive outcomes in offender programs (Andrews et al., 2006). In summary, offenders who are assessed to be at higher risk should be prioritized for intensive services, and these services should target criminogenic needs and responsivity factors in order to reduce recidivism and improve outcomes in treatment and supervision. Lower risk offenders do not require the same services or intensity of services to achieve comparable outcomes (Thanner & Taxman, 2003).

Risk-Needs-Responsivity (RNR) Simulation Tool

Crites & Taxman (2013) have developed a web-based Risk-Needs-Responsivity (RNR) Simulation Tool that categorizes community treatment programs according to their focus on evidence-based practices related to criminogenic needs and matches offenders to their particular level of risk and needs. The RNR Simulation Tool is based on the ASAM PPC model and a similar treatment matching model, Level of Care Utilization System (LOCUS), developed by the American Association of Community Psychiatrists (2009). The RNR Simulation Tool classifies offender programs by assessing several domains: target, content, dosage, and implementation quality. These domains are linked to increased effectiveness of offender programs (Andrews & Dowden, 2005). Information from each domain is then used to match offenders to specific programs. The following types of information are compiled for each domain:

- Target addresses the behavior(s) that are the focus of the particular treatment program. These include reducing the severity of substance use problems, cognitive restructuring of criminal thinking and reducing criminal peers, self-improvement and self-management strategies (e.g., improving social skills, problem solving, self-control), improving social/interpersonal skills, identifying deficits in physical/life needs (e.g., employment, education, housing), and implementing a sanctions-only approach for those who are at low risk. As noted previously, effective "targets" for offender treatment programs are those that address criminogenic needs that are linked to reducing recidivism (Andrews, 2012; Andrews & Bonta, 2010a, 2010b)
- Content addresses the therapeutic orientation of treatment programs, including the main area of treatment focus, services provided, and reinforcement of treatment skills. The content of offender programs should be a CBT skills-based approach to address factors such as antisocial behaviors, thinking, and peers, in addition to substance use disorders (Lipsey, Landenberger, & Wilson, 2007). Other key content includes social restrictiveness or supervision (e.g., curfews, probation visits, and mandatory daily program attendance), which can reduce recidivism (Drake, Aos, & Miller, 2009)
- Dosage addresses the amount (total number of hours), duration (number of weeks or months), frequency (number of times per week), and quantity (number of hours per week) of services provided by treatment programs. Dosage serves to moderate the risk for recidivism (Lipsey & Landenberger, 2005). Moreover, risk level determines the appropriate dosage necessary, with high-risk offenders generally requiring at least 300 hours of cognitive-behavioral treatment (CBT) and related services, moderate-risk offenders requiring approximately 200 hours of CBT and related services; and low-risk offenders requiring approximately 100 hours of services (Bourgon & Armstrong, 2005)
- Implementation Quality addresses whether programs are implemented as designed. Key factors include adherence to treatment protocols, proper staff training in delivering services, certification in administration of treatment protocols, supervision of staff who implement treatment protocols, use of quality assurance measures, and adequate

staff communication regarding participants' treatment progress

A second part of the RNR Simulation Tool involves profiling of offenders, based on offenders' risk level for recidivism. Risk level is composed of factors related to criminal history (leading to classification as "low," "moderate," or "high-risk" offenders), primary needs (e.g., substance use disorders, criminal thinking), clinical destabilizers (e.g., presence of mental disorders), lifestyle destabilizers (e.g., poor social supports, lack of education, unemployment, lack of stable housing), and stabilizers (i.e., opposite of destabilizing factors, such as educational achievement, stable housing, social support). Programs are categorized according to these features and placed in one of six groups (Crites & Taxman, 2013) that are differentiated by recidivism risk level, primary needs, responsivity (appropriate match between individual's needs and program services), dosage, program integrity (factors associated with implementation fidelity), and social restrictiveness.

Summary of Key Issues

- The Risk-Needs-Responsivity (RNR) Simulation Tool uses a series of algorithms generated from the Bureau of Justice Assistance, Survey of Inmates data set to match offenders with appropriate programs
- The tool also helps to identify gaps between offenders' needs and the existing program resources in a particular community (Crites & Taxman, 2013)
- The RNR model provides a useful framework to identify and address criminogenic needs and responsivity factors that influence treatment outcomes among offenders with CODs, including relapse and recidivism
- The RNR Simulation Tool is based on an empirically derived theoretical approach to identify the appropriate level of treatment and supervision services that are needed to promote positive outcomes among offenders who have substance use problems and CODs

Concerns

- Although the RNR Simulation Tool is based on a sound theoretical model to determine treatment matching for those involved in the justice system, it is a new approach and requires application and testing to assess its validity, including its effectiveness in reducing recidivism
- Several other assessment tools are available to examine offenders' risk and needs for psychosocial interventions. These include the Addiction Severity Index (ASI; McLellan et al., 1985), the Global Assessment of Individual Needs (Dennis, Titus, White, Unsicker, & Hodgkins, 2003), the Level of Service Inventory-Revised (Andrews & Bonta, 1995), and a range of other risk assessment instruments

Availability and Cost

Information regarding the RNR Simulation Tool is available at the following site: http://www.gmuace.org/tools/. Direct link to the RNR Simulation Tool: http://www.gmuace.org/tools/program-tool

American Society of Addiction Medicine-Patient Placement Criteria (ASAM PPC)

The ASAM PPC is a widely used assessment and triage approach that employs patient placement criteria to identify appropriate levels of care for people who have substance use disorders and CODs. The ASAM PPC for the Treatment of Psychoactive Substance Use Disorders (Hoffman, Halikas, Mee-Lee, & Weedman, 1991) were developed through a consensus process, and this approach has subsequently been used in a number of states and increasingly by managed care organizations to modify treatment matching approaches for use in the behavioral health field. The ASAM PPC were revised in 1996 and again in 2001 (ASAM PPC-2R; Mee-Lee, Shulman, Fishman, Gastfriend, & Griffith, 2001). The most recent revision, ASAM Criteria-Treatment Criteria for Addictive, Substance Related, and Co-occurring Conditions (Mee-Lee, 2013), reflects changes to the DSM-5 diagnostic criteria.

Underlying concepts of the ASAM PPC (Mee-Lee & Shulman, 2003) include the following: (1) the biopsychosocial perspective of addiction that encompasses etiology, expression, and treatment of addiction, allowing for a more comprehensive assessment and treatment approach; (2) individualized treatment that provides a patient-driven approach; (3) multidimensional assessment (see the six domains below) that determines level of services needed; (4) treatment matching that integrates all six domains (described in the following section) and addresses issues of motivation to change, management of social/occupational risk factors, medication management (e.g., detoxification, craving management), and other services (e.g., self-help/12-step groups, such as NA and Dual Recovery Anonymous); and (5) monitoring of care that includes relapse prevention, treatment engagement and retention, and other important social/occupational factors.

The ASAM PPC provide separate guidelines for placement in adolescent and adult treatment services. The ASAM PPC-2R guidelines operationalize six assessment dimensions that define biopsychosocial severity within the context of behavioral health services: (1) acute intoxication and/or withdrawal potential; (2) biomedical conditions and complications; (3) emotional, behavioral, or cognitive conditions and complications; (4) readiness to change; (5) relapse, continued use, or continued problem potential; and (6) recovery/living environment. Criteria described for each of the six dimensions are then used to guide placement in one of five levels of treatment services, which vary by the intensity of services provided: (1) level 0.5—Early intervention, (2) level I—Outpatient treatment, (3) level II—Intensive outpatient/partial hospitalization treatment, (4) level III—Residential/inpatient treatment, and (5) level IV—Medically managed intensive inpatient treatment.

The ASAM PPC-2R (2001) were the first to identify the need for substance use programs to provide integrated services for CODs. The ASAM PPC-2R supplement also reviews issues related to medically assisted treatment for alcohol use disorders (AUDs), detoxification, and relapse prevention. The ASAM PPC-2R guidelines recognize that for people with CODs, whichever disorder causes the most functional impairment should be considered in making the placement to a particular type of treatment setting. Treatment programs described in the PPC-2R may be either "dual diagnosis capable" or "dual diagnosis enhanced," to address people with CODs who demonstrate a wide range of psychopathology. Specifically, dual-diagnosis capable programs are those that address the comorbidity between substance use disorders and more stable mental health problems, where the co-occurring mental health problems do not interfere with engagement and progress in addiction treatment. Policies and procedures address dual diagnoses and allow for collaboration with mental health services to appropriately handle CODs and provide psychopharmacological monitoring/assessment both on site and via coordinated off-site services. Dual diagnosis enhanced programs accept individuals who have CODs and more unstable mental disorders. These programs allow for mental health problems to be managed simultaneously with addictions, providing continuity in the overall treatment approach. Policies and procedures include more stringent monitoring of participants and integration of mental health treatment with addictions treatment, which allows for treatment continuity for both disorders. For each level of treatment, criteria are specified (within dimensions 2–6) for dual-diagnosis capable and enhanced programs.

ASAM developers provide a range of information to aid in standardizing clinical assessment and placement, in addition to materials to encourage individualized treatment planning. Tutorials and distance learning are also provided to help train individuals in proper assessment and appropriate treatment placement. The instrument also employs automated software that utilizes an algorithm (Turner, Turner, Reif, Gutowksi, & Gastfriend, 1999) for matching individuals with appropriate treatment programs. This software application

demonstrates good concurrent validity with other standardized assessments, such as the Addiction Severity Index (ASI), and predicts treatment outcomes for those who are appropriately matched (Magura et al., 2003; Sharon et al., 2003).

One caveat to these research findings is that many individuals were mismatched for treatment or did not show up to treatment and thus were not included in these results (Angarita et al., 2007; Gastfriend & Mee-Lee, 2011). In a study of alcohol users, those who were mismatched to more intensive levels of treatment did not show greater improvement in treatment outcomes than those who were correctly matched to treatment. However, people mismatched to less intensive levels of treatment showed poorer treatment outcomes (Magura et al., 2003). Another study indicated that those who needed higher levels of care did not receive it (e.g., residential treatment Level III versus hospitalization Level IV) and were in treatment significantly longer than those who were matched to the correct level of care (Sharon et al., 2003).

Difficulty in treatment matching may be due in part to substantial disagreement (81 percent) between computerized algorithm results and clinician recommendations (Sharon et al., 2003). Clinicians may judge the algorithm's matching recommendations as too restrictive. The algorithm may classify individuals into higher levels of treatment based on one item in the PPC criteria rather than considering other items that provide more relevant coverage of that particular dimension. For example, concerns related to emotion/behavioral functioning may lead to matching people to Level IV, but these people may be just as well suited as people matched to Level III to complete the treatment program successfully.

Challenges in Applying the ASAM Criteria in Justice Settings

Although the ASAM criteria have been commonly used in community-based settings to guide treatment matching, they have only recently been implemented in the justice system. For example, only about a third of drug court survey respondents indicated the use of the ASAM PPC (American University, 2001). Several states now use the ASAM criteria to place individuals convicted of DUI/DWI offenses in different types of treatment programs. The ASAM PPC or similar approaches provide a structured approach to potentially match justice-involved individuals more effectively to different levels of treatment intensity, structure, and supervision (CSAT, 2005b).

There are several challenges in implementing the ASAM criteria in justice settings (Mee-Lee, 2013). First, specific to readiness to change, there may be an unreasonable expectation, particularly in the first few months of treatment, that offenders are in the "action stage" of recovery and are able to comply with justice system mandates for abstinence from drugs and alcohol and fully engage with treatment services. In addition, some treatment programs that are mandated by the courts may be too short in duration for participants to reach the "action stage" of recovery and to maintain healthy and prosocial behaviors.

Some judges or community supervision officers may also place offenders in mandated treatment based on their own view of what level of care is needed rather than by conducting a formal assessment to identify treatment needs and match people to appropriate services. In contrast, some courts may recommend treatments that seem more "restrictive" such as residential programs, in part because the proxy of confinement gives a sense of comfort related to criminal recidivism potential or violence risk reduction. This can be problematic if the treatment needs are not as intensive as the treatment that falls under a court order.

In other justice settings, offenders are placed in treatment based on the resources that are available rather than on individualized needs for treatment. As a result, offenders may not receive a comprehensive assessment or the optimal services that are needed. Another consideration is that the recent emphasis on risk assessment procedures in justice settings may result in offenders receiving

treatment and supervision that is focused primarily on antisocial behaviors, attitudes, and peers, without considering the importance of other factors, such as co-occurring mental disorders and substance use issues, employment, education, and family services, that also influence criminal involvement and recovery.

Finally, the ASAM PPC are based on a medical model of substance use treatment that includes an emphasis on individual counseling and oversight provided by medical personnel, whereas group counseling is the preferred approach for offenders (including those with substance use disorders), and oversight is typically provided by justice or substance use treatment personnel. A related concern is that the ASAM PPC do not currently provide a "dimension" that addresses risk for criminal recidivism, nor does the PPC provide recommendations for how to modify "levels" of treatment to address the unique resources and limitations related to drug courts, day treatment, other community correctional treatment programs, or jail and prison-based programs.

Summary of Key Issues

- Implementation of the ASAM PPC-2R criteria includes the use of standardized assessment tools and computerized software, which can improve accuracy in matching individuals to appropriate treatment programs (Baker & Gastfriend, 2003; Gastfriend & Mee-Lee, 2011)
- A study involving outpatient treatment programs provides support for the ASAM model in treatment matching and indicates that programs using standardized ASAM PPC assessment tools are more likely to provide both counseling and other support services that follow practice guidelines developed by ASAM or CSAT (Rieckmann, Fuller, Saedi, & McCarty, 2010)
- The Addiction Severity Index (ASI) is a common standardized assessment tool used in ASAM implementation in outpatient settings and criminal justice settings (Cohen, Mankey, & Wendt, 2003; Koob, Brocato, & Kleinpeter, 2011; Magura et al., 2003; Marlowe, Festinger, Dugosh, Arabia, & Kirby, 2008; Rieckmann et al., 2010)
- The Global Assessment of Functioning (GAF) and Structured Clinical Interview for Diagnostic Statistical Manual (SCID) are commonly used for mental health assessment and diagnosis in treatment settings that use the ASAM criteria (Kosanke, Magura, Staines, Foote, & DeLuca, 2002; Magura et al., 2003; Rieckmann et al., 2010)

Concerns

- Challenges in implementing the ASAM PPC criteria in justice settings include the need to address criminal risk as it affects placement in various levels of treatment and supervision, matching to specialized offender programs (e.g., drug courts), the need to triage offenders to programs that provide group treatment services, and the need to integrate specialized CODs treatment services with intensive supervision and court monitoring
- Further research is needed to establish the validity of the ASAM PPC in improving treatment outcomes among offenders who have substance use disorders and CODs
- Although the ASAM PPC computerized software helps to predict treatment outcomes among people matched to various levels of treatment, studies examining placement outcomes using the ASAM PPC criteria generally do not include people who were mismatched to treatment and who did not attend treatment. Many individuals who are mismatched to treatment show poorer treatment outcomes. In addition, there is significant disagreement between ASAM PPC treatment placements generated by the computerized algorithm and clinician-recommended treatment placements. It is important to consider factors that may contribute to these disparities, including the emphasis placed on certain dimensional criteria by the computerized algorithm.

Further research is needed to examine treatment outcomes among people who are mismatched to treatment based on the ASAM PPC computerized algorithm, and to identify strategies to reduce these mismatches

- The ASAM PPC materials are somewhat costly to purchase

Availability and Cost

The most recent version of the ASAM PPC, The ASAM Criteria: Treatment Criteria for Addictive, Substance-Related, and Co-occurring Conditions and the ASAM PPC supplement can be purchased from the American Society of Addiction Medicine at the following site: http://www.asam.org/publications/the-asam-criteria

The cost of the ASAM PPC is $95 ($85 for members of ASAM), and the supplement costs $65 and is available for the Kindle.

ASAM recommends a set of assessment and placement instruments that adhere to ASAM criteria, and these are available for purchase. Assessment and placement instruments cost between $50 and $80, and each instrument contains 25 copies. Instruments can be obtained at the following site: http://changecompanies.net/asamcriteria/assessments.php

Substance Use Assessment Instruments and Treatment Matching

Several assessment instruments have been developed for treatment matching as part of the RNR Simulation Model and the ASAM PPC, as described in previous sections. A number of risk assessment instruments are also available to assist in matching to treatment and supervision, as described in the "Risk Assessment" section of this monograph. Several other substance use assessment instruments are frequently used in treatment matching in behavioral health settings and are described in this section. These include the Addiction Severity Index (ASI), the Texas Christian University intake and assessment forms/instruments, and the Timeline Followback (TLFB).

Addiction Severity Index-Fifth Version (ASI-5/ASI-6)

The ASI (McLellan et al., 1992; McLellan, Luborsky, Woody, & O'Brien, 1980) is one of the most widely used instruments for screening, assessment, and treatment planning related to substance use disorders. The 155-item instrument was designed as a structured interview to examine symptoms, frequency of substance use, and other psychosocial areas that are frequently affected by substance use. The ASI requires 45–60 minutes to administer and 10–20 minutes to score. Additional versions of the instrument have been developed for clinical and training purposes (ASI-CTV), and a brief version is available that takes approximately 30 minutes to administer (ASI-Lite). The ASI-Lite has been adapted for use in the VA system (ASI-L-VA).

Self-report and clinician administered computerized versions of the ASI are available (ASI-Net and CA ASI-Net), as are versions designed for interactive voice response (ASI-IVR) and automated telephone administration (Brodey et al., 2004; Rosen et al., 2000). The ASI-Multimedia Version (ASI-MV; Butler et al., 2001) is a computerized form of the instrument, and was designed to reduce burden on treatment counselors. The instrument provides virtual simulation of a clinician-administered interview and includes audio and video presentations as well as "skip-logic." The instrument has been found to be reliable and valid (Butler et al., 2001) and generates two summary scores: (1) composite scores for each ASI domain, and (2) severity ratings by domain for problems occurring during the past month. The composite scores generated by the interview and automated versions of the ASI are highly correlated (.91), indicating high reliability between the different versions of the instrument (Brodey et al., 2004).

The ASI includes seven domains of functioning commonly affected by substance use, including drug and alcohol use (separate sections), legal status, family and social relationships,

employment and support status, medical status, and psychiatric status. The ASI examines the severity of problems in each of these domains over the past month and the need for treatment. The instrument also reviews indicators of emotional, physical, and sexual abuse. Although the ASI measures frequency of use, it does not address quantity of use, as quantity may be underestimated and frequency is easier to recall (McLellan et al., 1992). The ASI-5 includes interviewer severity ratings (ISR) that combine current and lifetime symptoms within each domain to help assess the need for treatment. The ASI composite summary scores (CS) are generated for each domain and assess the current severity of symptoms. Evaluation factors (EF) are available for five of the domains, and clinical factors (CF) are included for all seven domains. CFs measure current and lifetime functioning scores that reflect a global severity rating. EFs measure individual functioning during the past month.

Many offender programs have developed modified versions of the ASI for use in substance use screening. A sixth edition of the ASI is now available. Revisions to the ASI-6 include replacement of the ISR ratings with clinical indices of lifetime functioning (CIs). An interval of 6 months has been added in addition to past month and lifetime ratings. The ASI-6 includes "skip-out" questions that can reduce administration time to approximately 1 hour, and the instrument provides more specific wording of questions to increase reliability. Item Response Theory (IRT) analysis indicates that in comparison to previous versions, the ASI-6 is better able to address changes in substance use problems and treatment needs of diverse populations (e.g., welfare clients, drug court participants, individuals who are homeless) and has improved psychometric properties across the seven domains. The ASI-6 consists of nine summary scores ("Recent Status Scores" or RSS) that map to the seven Composite Scores in the ASI-5, with two additional summary scores that address family/social support and child problems (McLellan, Cacciola, Alterman, Rikoon, Carise, 2006; Denis, Cacciola, Alterman, 2013).

The ASI-6 also contains a follow-up interview that addresses change in symptoms over time. Items from the ASI-6 differentiate between current symptoms (past 30 days) and those experienced since the last administration of the ASI interview.

Positive Features

- The ASI-6 has been translated into Spanish and several other languages
- The ASI is a public domain instrument and is available at no cost
- The ASI describes recent and long-term patterns of substance use and examines a range of different legal and illegal substances. The ASI can also be used to screen for trauma and PTSD (Cacciola et al., 2007; Najavits et al., 1998). The ASI-6 provides more structure than previous versions of the instrument and enhanced ability to identify drug, alcohol, and mental health problems (Cacciola, Alterman, Habing, & McLellan, 2011)
- Recent validity studies indicate improvement of several scales on the ASI-6 in comparison to the ASI-5 (Denis et al., 2013)
- Many criminal justice agencies have used sections of the ASI-6 for substance use screening (McLellan et al., 1985; Peters et al., 2000), as well as the full ASI-6 for assessment purposes (Eriksson et al., 2013; Ettner et al., 2006; Pankow et al., 2012; Proctor, 2012; Serowik & Yanos, 2013)
- Among offenders, the ASI-6 (McLellan et al., 2006) shows good concurrent validity, including significant correlations with the Texas Christian University Drug Screen II (TCUDS-II), a validated substance use screening measure. Scores from the ASI-6 domains are significantly correlated with scales from other TCU instruments. For example, the ASI-6 is significantly correlated with the TCU psychological functioning (PSYForm)–self-esteem scale; the TCU social functioning (SOCForm) scales of social desirability, social functioning, and hostility; the

psychological functioning scales of anxiety/depression; and the TCU criminal thinking scales (CTS; Pankow et al., 2012). The ASI-6 (Pankow et al., 2012) is also significantly correlated with other validated psychological measures, such as the K10 (Kessler et al., 2003) and the PTSD Checklist (PCL; Weathers et al., 1993)

- ASI normative data is available for criminal justice populations (McLellan et al., 1992)
- The ASI is highly correlated with objective indicators of addiction severity (McLellan et al., 1980, 1985; Searles et al., 1990) and with alcohol use disorder and substance use disorder diagnoses (Rikoon, Cacciola, Carise, Alterman, & McLellan, 2006). The ASI-Drug Use section was one of three sets of screening instruments found to be the most effective in identifying substance-dependent offenders (Peters et al., 2000)
- Among people seeking substance use treatment, the ASI-6 domains/scales show good concurrent validity with other related measures and are correlated with measures of the following: (1) medical problems and physical health, as measured by the Short Form Mental Health Survey (SF-12, r score = -.64); (2) family/social support, as measured by the Social Readjustment Scale Self-Report, SAS-SR-social (r score = -.34); (3) family and social problems, as measured by the SAS-SR social (r score = .40); (4) employment, as measured by the SAS-SR Work, (r score = .76), (5) alcohol problems, as measured by the Short Index of Problems (SIP-Alcohol, r score = .68; Alterman, Cacciola, Ivey, Habing, & Lynch, 2009); (6) drug problems, as measured by the SIP-Drugs (r score = .61; Alterman et al., 2009); (7) legal problems, as measured by prior arrests (r score = .15); and (8) mental health problems, as measured by the Symptom Checklist Revised (SCL-10R, r score = .68; Cacciola Alterman, Habing, & McLellan, 2011)
- Among people with substance use disorders, the ASI-5 domains/scales also demonstrate good concurrent validity with other related measures of physical health, current/lifetime alcohol problems, recent/lifetime drug problems, legal problems, and family/social problems (Alterman et al., 2009). The ASI-6 domains may provide better coverage than the original ASI-5 domains, particularly the family/social area and its subscales (Denis et al., 2013). The ASI-6 also demonstrates higher correlations than the ASI-5 with concurrent validity measures in five of the seven original domains (employment, psychiatric, family/social, legal, and drug; Denis et al., 2013)
- The ASI-6 has good internal consistency across all domains, the summary scales, and across different race/ethnicity groups (alphas range .73–.94; Cacciola et al., 2011). Most of the ASI-6 RSS domains are highly correlated with the ASI-5 CS scales (Denis et al., 2013)
- When administered over a 2–3 day period to a treatment-seeking sample, the ASI-5 has good interrater reliability for agreement with the ASI-L-VA on most ISR ratings and scores for CS, CF, and EF, across domains of alcohol, drugs, and psychiatric problems (ICCs range .62–.89; Cacciola et al., 2007). Similarly, the ASI-5 has adequate test-retest reliability for most ISRs ratings and CS, CF, and EF scores, when readministered after short intervals (Cacciola et al., 2007)
- The seven domains of the ASI-5 have good internal consistency (alphas range .73–.92) for both current and lifetime problems (Alterman, Cacciola, Habing, & Lynch, 2007)
- The ASI-5 has acceptable internal consistency for most summary scales (CFs, CSs, Efs; alphas range .72–.89; Cacciola et al., 2007). The ASI-L-VA has acceptable internal consistency across the same summary scales (Cacciola et al., 2007)
- Research indicates that the ASI is reliable and valid for use with people who have CODs (Carey, 1997)
- In comparison to the ASI-MV, the ASI-5 demonstrated no significant differences

in responses for particular domains such as employment, and items specific to alcohol use (Butler, Villapiano, & Malinow, 2009). Areas of significant differences that were found could be due to higher rates of disclosure by participants on the computerized interview as compared to face-to-face interviews (Butler et al., 2009; Garb, 2007)

Concerns

- The ASI-6 is still in the process of development and is not as widely used as the ASI-5
- The ASI requires approximately 45–90 minutes to administer, although the alcohol and drug sections can be completed in significantly less time
- Substantial training is needed to administer and score the ASI
- The ASI-6 Spanish version demonstrates variable psychometric properties, including poor to good internal consistency (alphas range .47–.95; Díaz-Mesa et al., 2010) and poor to excellent test-retest reliability (.36–1.0; Díaz-Mesa et al., 2010)
- The ASI-5 legal scales may be more valid than those of the ASI-6 (Denis et al., 2013). For example, ASI-5 arrest results from the ASI-5 legal domain are more highly correlated with the history of arrest than the ASI-6 (Denis et al., 2013)
- Among those seeking substance use treatment, the ASI-5 has lower interrater reliability for agreement ISR ratings in domains of employment and family-social problems and lower EFs, CFs, and CSs for family-social problems when compared to the ASI-L-VA (ICCs < .60; Cacciola et al., 2007), indicating that these domains may generate inconsistent or inaccurate ratings
- The ASI-5 may have poor test-retest reliability for EF, CS, and ISR ratings related to the family/social domain (ICCs < .60; Cacciola et al., 2007)
- The ASI-5 may have lower internal consistency for certain summary scales, such as drug (CS) and legal problems (CS/CF; alphas <.70; Cacciola et al., 2007). The ASI-L-VA also exhibits lower internal consistency on these scales (Cacciola et al., 2007)
- Results from the ASI-MV (Butler et al., 2001) and face-to-face interview versions of the ASI may be inconsistent, as differences in scores were obtained in the following domains: drug, alcohol, legal, family, and psychiatric problems (Butler et al., 2009)
- The ASI may have reduced reliability and validity for people who have significant substance use problems and co-occurring mental disorders (Carey, 1997; Corse, Hirschinger, & Zanis, 1995; McLellan, Cacciola, & Alterman, 2004; Zanis, McLellan, & Corse, 1997)

Availability and Cost

The ASI is a public domain instrument that was developed by the Treatment Research Institute, 600 Public Ledger Building, 150 South Independence Mall West, Philadelphia, PA 19106, (215) 399-0980. The instrument is available at the following site: http://www.tresearch.org/index.php/tools/download-asi-instruments-manuals/

This site also provides several manuals that include information on administration, scoring, and interpreting the ASI.

The ASI-6 is available at no charge on a case-by-case basis. Additional information regarding the ASI-6 can be obtained by emailing the help desk at ASIHelpline@tresearch.org

Texas Christian University (TCU) Intake and Assessment Instruments

The TCU intake and assessment instruments (Simpson & Knight, 1998) are available in the public domain and include versions tailored specifically for criminal justice and community treatment settings. The instruments assess a broad range of psychosocial issues, including drug, alcohol, and mental health problems, as well as

other social, occupational, and treatment areas. TCU instruments described here include both interviewer administered and self-report scales. Instruments developed for justice settings are referred to as the Criminal Justice treatment forms (TCU CJ) and contain an interviewer-administered CJ Comprehensive Intake (TCU CJ CI), and a self-report CJ Client Evaluation of Self and Treatment (TCU CJ CEST-intake). Instruments developed for community treatment settings include an interviewer-administered Brief Intake (TCU BI), a Comprehensive Intake (TCU CI), and a self-report Client Evaluation of Self and Treatment, Intake version (TCU CEST-Intake).

The self-report CEST forms for both criminal justice and community settings contain several sections, or short forms, that can be administered separately. A follow-up CEST form is also available for both community and justice settings and can be used to evaluate treatment progress over time. Other self-report instruments can be combined with both the criminal justice and community CEST forms, including the TCU Drug Screen V (TCUDS V), the TCU Criminal Thinking Scales (TCU CTS), and other mental health scales that integrate components of the K6 and K10 instruments (Kessler et al., 2003). Several TCU short forms are based on sections contained in the original interviewer-administered intake instrument. These include the global risk assessment (TCU RSKForm), the Family and Friends assessment (TCU FMFRForm), the mental health and PTSD screen (TCU TRMAForm), and physical and mental health screens (TCU HTLHForm). The TCU HTLHForm contains items from the K10 and is designed to examine psychological distress. The short forms provide a vehicle for individualized assessment to address CODs relevant to involvement in treatment.

Criminal Justice Instruments:

- The TCU CJ Comprehensive Intake (TCU CJ CI) is administered 1 to 3 weeks after program entry and queries about the past month or the past 6 months prior to incarceration. The TCU CJ CI contains sections assessing the following domains:
  - Sociodemographic background
  - Family background, including quality of relationships with family members
  - Peer relations, including quality of relationships with friends and gang affiliations
  - Criminal history, including prior arrests, involvement in illegal activities, and legal status
  - Health and psychological status, including physical and mental health (e.g., anxiety, depression), and history of hospitalization
  - Drug history, including frequency of alcohol and drug use over the past month and past 6 months and prior treatment history. Alcohol use is assessed in more detail, including quantity and patterns of drinking over the past month. Problems caused by drug and alcohol use are based on DSM-IV criteria
  - AIDS risk assessment, including risky behaviors

  The TCU CJ CI requires approximately 90 minutes to administer. Instructions are provided to the interviewer to read aloud to the participant explaining the purpose of the assessment, in addition to answer cards to help guide the format of participants' responses. "Skip logic" items are provided that can reduce the duration of administration.

- The TCU CJ Client Evaluation of Self and Treatment (TCU CJ CEST; Joe, Broome, Rowan-Szal, & Simpson, 2002; Knight, Simpson, & Morey, 2002) is a self-report instrument for use with offenders. The instrument examines treatment motivation and a range of other psychosocial factors affecting treatment. The TCU CJ CEST reviews the following domains:
  - Treatment motivation, with subscales of problem recognition (PR), desire for

help (DH), treatment readiness (TR), and pressure for treatment index (PT)

- » Psychological functioning, with subscales of self-esteem (SE), depression (DP), anxiety (AX), and decision making (DM)
- » Social functioning, with subscales of childhood problems (CP), hostility (HS), and risk taking (RT)

A scoring guide is provided to help interpret results from the instrument. Each of the TCU CJ CEST domains can be administered as separate one-page forms, in combination with each other, with other scales (TCU CTS, TCUDS V, K6/K10), or with other short forms, as described previously, to provide a more individualized assessment approach. The short forms and scales are designed to supplement intake assessments that are used by different justice programs. Individual scoring manuals are provided for each of the short forms. The follow-up version of the CEST also contains a "treatment progress domain" that provides subscales related to treatment participation (TP), treatment satisfaction (TP), counseling rapport (CR), peer support (PS), and social support (SS). The treatment progress domain can also be administered as a separate one-page form. A follow-up version of the CEST can be administered over the course of treatment to assess change over time for each of the domains and to examine engagement and retention, as indicated by the treatment progress domain.

Community Treatment Forms:

The TCU community treatment instruments are similar to the criminal justice instruments but are designed primarily for outpatient and residential treatment settings.

- **The Brief Intake interview (TCU BI)** contains sections similar to the CJ Comprehensive Intake but is significantly shorter. The instrument includes the following sections:
  - » Background information
  - » Psychosocial functioning during the past 6 months
  - » Drug use background, including information describing substance use in the past 6 months and during the lifetime
  - » Drug use problems in the past year, including areas addressed by the DSM criteria for substance use disorders
- **The Comprehensive Intake Interview (TCU CI)** is similar to the TCU CJ CI interview but is geared towards those receiving treatment in the community and includes special instructions for those entering treatment from jail or prison. Domains of the TCU CI are similar to those in the TCU CJ CI, but there are several differences in the item structure and wording of individual items. For example, the sociodemographic background section provides detailed information about childhood history. The drug history section includes questions addressing treatment support from family and friends and problems related to gambling. An additional section is provided to record interviewer comments about the quality of participant responses. The TCU CI requires approximately 90 minutes to administer, and like the TCU CJ CI, includes answer cards and instructions for administration.
- **The Client Evaluation of Self and Treatment Intake Version (TCU CEST-Intake)** is a self-report instrument that is similar to the TCU CJ CEST and that includes similar domains addressing treatment motivation, psychological functioning, treatment motivation, and social functioning. As with the TCU CJ CEST, each domain of the TCU CEST-Intake can function as a stand-alone instrument or be combined with other short forms. Unique to the TCU CEST-Intake is a self-efficacy scale (Pearlin Mastery Scale (Pearlin & Schooler, 1978) that is embedded in the psychological functioning domain. A social consciousness scale is

also included in the social functioning domain and examines social values. The follow-up CEST-Intake is identical to the CJ CEST version in coverage of domains and analysis of treatment engagement, retention, and progress. A manual is provided to assist in scoring and interpretation of the CEST-Intake.

Positive Features

- The TCU intake and assessment instruments have been used in a wide variety of offender settings (Farabee, Prendergast, & Cartier, 2002; Czuchry & Dansereau, 2000; Joe, Rowan-Szal, Greener, Simpson, & Vance, 2010; Pankow & Knight, 2012)
- The TCU CJ CEST and community CEST instruments include norms for both offender and community treatment populations
- The TCU intake and assessment instruments provide two sets of forms that are tailored for offender and community treatment settings
- Each of the TCU intake and assessment instruments is fully structured and addresses multiple domains, including diagnostic criteria for various disorders. The instruments can be administered by nonclinicians and include a straightforward set of items/questions
- The self-report CEST forms can be administered as short, one-page assessments or can be combined to provide a more comprehensive assessment, thus allowing programs flexibility to tailor their approach to the needs of participants and to the needs of the program. For example, several short forms are available to assess mental health, social functioning and other related domains, and these can be administered individually or in combination with CEST forms
- The assessment forms examine DSM criteria for drug and alcohol use disorders. The self-report TCU CJ CEST can be combined with other forms, such as the TCU CTS, to assess risk for recidivism and to provide a more comprehensive assessment. Criminal thinking as measured by the TCU CTS is correlated with lower treatment motivation/engagement and poorer psychological and social functioning (Garner et al., 2007)
- TCU CJ CEST motivation scales are correlated with treatment engagement among offenders (Pankow et al., 2012; Simpson et al., 2012)
- The TCU CJ CEST domains of psychological functioning, social functioning, and motivation are related to relevant domains on the Addiction Severity Index, supporting the convergent validity of the CEST instrument. For example, treatment motivation and psychological and social functioning are correlated with ASI measures of legal status, drug problems, and psychiatric problems (Pankow et al., 2012)
- Among female offenders, the TCU TRMAForm and TCU HLTHForm are highly correlated with the psychological functioning scales/domains of anxiety and depression in addition to social functioning scales/domains of hostility and risk taking, supporting the concurrent validity of these measures (Rowan-Szal et al., 2012)
- The TCU CJ CEST shows acceptable internal consistency in justice settings across domains of treatment motivation (alphas range .60–.80), psychological functioning (alphas range .71–.74), and social functioning (alphas range .71–.80; Garner et al., 2007). Other studies provide support for the internal consistency of the entire CEST instrument (Simpson, 2004; Simpson, Knight, Dansereau, 2004) and for the specific domains that can be used as independent assessment instruments (e.g., TCU psychological functioning and TCU social functioning domains; Rowan-Szal et al., 2012; Simpson et al., 2012)
- TCU CJ CEST subscales of social functioning and psychological functioning represent unitary dimensions, as indicated

by confirmatory factor analyses (Garner et al., 2007; Simpson et al., 2012)

- The CJ CEST domains have good test-retest reliability across subscales (Garner et al., 2007)
- The TCU CEST Community Treatment forms demonstrate good internal consistency for domains of treatment motivation (alphas range .88–.90), social functioning (.71–.90), and psychological functioning (.80–.91; Joe et al., 2002; Simpson, 2004)
- The TCU TRMAForm, TCU HLTHForm, and their subscales show good internal consistency among female offenders (alphas range .75–.94; Rowan-Szal et al., 2012)

Concerns

- Further study is needed to determine the validity and reliability of both the TCU intake and assessment forms in detecting the severity and scope of substance use disorders, mental disorders, and related psychosocial problems
- Many of the existing studies of the TCU intake and assessment forms in justice settings have been conducted by the developers of the instruments. Studies conducted by other research teams are needed to confirm these results
- The criminal justice and community treatment intake and assessment forms do not include a module to detect psychosis
- The TCU CEST does not address antisocial behaviors
- The domain of treatment motivation and its subscales appear to have relatively low internal consistency, particularly the subscales related to desire for help (alpha = .67) and treatment needs (alpha = .60). Results of confirmatory factors analyses indicate that the four treatment motivation subscales may lack structural integrity and may not represent unitary dimensions (Garner et al., 2007)

Availability and Cost

Each of the TCU intake and assessment instruments is available at no cost. The community treatment forms, including scoring interpretation and norms can be found at the following site: http://ibr.tcu.edu/forms/tcu-core-forms/

Criminal justice treatment forms including scoring, interpretation and norms can be found at the following sites:

http://ibr.tcu.edu/forms/forms-archives/cj-forms-correctional-residential-treatment/

http://ibr.tcu.edu/forms/forms-archives/cj-forms-correctional-outpatient-treatment/

The individual CEST domains as one-page forms and a scoring guide for the implementation of the CEST can be obtained from the following site: http://ibr.tcu.edu/forms/client-evaluation-of-self-and-treatment-cest/

Other TCU forms can be found at the following site, which links the user to archives containing various forms and descriptions of each form: http://ibr.tcu.edu/forms/forms-archives/

Timeline Followback (TLFB)

The Timeline Followback (TLFB) protocol provides a detailed daily history of alcohol and other substance use over a specific period of time (from 7 days to 2 years) but is employed most frequently to examine substance use within the previous 3 months. The TLFB involves using a blank calendar to help produce a detailed pattern of substance use and nicotine use over specified time intervals. The calendar is used to help individuals identify and note memorable occasions over these time intervals (e.g., the past 30 days) to aid in the recall of daily patterns of substance use and nicotine use. Common variables computed for alcohol use include the number of drinking days, average drinks, total drinks per month, and maximum drinks consumed during one occasion (Pedersen & LaBrie, 2006). For drug

use, variables calculated include the number of days of use, the longest period of use, and the longest period of abstinence; however, this varies across drug class. For example, the quantity of marijuana use can be more accurately assessed in terms of frequency (number of joints; Robinson, Sobell, Sobell, & Leo, 2012). The TLFB approach provides a more accurate and comprehensive assessment of individual drinking and drug use patterns compared with typical quantity and frequency measures that may underestimate substance use behavior (Sobell et al., 2003). The TLFB protocol requires approximately 10–30 minutes to complete and is available in several languages.

Positive Features

- The TLFB measure can be administered via interview or computer. The computerized version provides detailed instructions for self-administration and allows measurement of time intervals up to 12 months. The computerized version of the TLFB requires the same amount of time to administer as the interview version
- The TLFB has been used successfully with justice populations (Broner, Mayrl, & Landsberg, 2005; Easton et al., 2007), including DUI/DWI offenders (Brown et al., 2008; Fridell, Hesse, & Billsten, 2007; Palmer, Ball, Rounsaville & O'Malley, 2007)
- In a meta-analysis of drug-involved populations (Hjorthøj, Hjorthøj, & Nordentoft, 2012), agreement between biological assessment (e.g., urine drug tests) and the self-report TLFB is quite good across drug classes (79–94 percent). Agreement between biological measures and the self-report TLFB is quite good across different time periods assessed by the TLFB. For example, with a period of less than 30 days, TLFB agreement ranges 81–85 percent, and for over 30 days, agreement ranges 87–93 percent. The TLFB produces few false negative errors for most categories of drugs when compared to urinalysis (Westerberg, Tonigan, & Miller, 1998)
- In comparing biological assays and the TLFB for specific drug classes during the past 60 days, agreement was 86–92 percent for cocaine and 84–87 percent for cannabis (Stasiewicz et al., 2008). Agreement across multiple substances during the past 6 months is also high (kappas = .74–.94; Morgenstern, Hogue, Dauber, Dasaro, & McKay, 2009), providing support for the reliability and validity of the TLFB over time
- Comparisons between the TLFB and ASI for people with CODs have found high rates of agreement between the two instruments (kappa = .79; Carey, 1997). However, the TLFB may yield higher estimates of drinking than the ASI over a 30-day interval
- In support of the concurrent validity of the TLFB among those enrolled in residential substance use treatment, the TLFB shows adequate agreement with the ASI (past 30 days) for reported alcohol use among people with substance use disorders (SUDs) only and for people who have CODs (91–93 percent agreement, kappas range .60–.70). Agreement is also high for drug use (82–87 percent, kappas range .63–.70; DeMarce et al., 2007)
- For samples with either SUDs or CODs, the TLFB demonstrates good agreement with collateral reports of alcohol use (90-91 percent; kappas range .50–.61) and with drug use (77–81 percent: kappas range .45–.62). Good agreement was also found between the TLFB and frequency of drinking days, as measured by the ASI (r scores range .70–.78) and collateral reports (r scores range .52–.62) in both samples (DeMarce et al., 2007)
- The TLFB is highly correlated with self-report measures of drug use frequency (DUF) over the previous 6 months across all drug classes (O'Farrell, Fals-Stewart, Murphy, & Murphy, 2003; r scores range .83–.96). Very high rates of agreement

have also been found between the TLFB and DUF on use versus non-use across all drug classes (r scores range .97–1.00; O'Farrell et al., 2003)

- The TLFB is highly correlated with measures of general life functioning (r scores = .62–.99; Westerberg et al., 1998)
- The test-retest reliability of the TLFB over 1–2 weeks is quite good among people with substance use disorders seeking treatment, for percent of days abstinent, longest period of use, and longest period of abstinence over 30, 60, and 360 days, for both cocaine (ICCs range .74–.90; r scores range .75–.91) and marijuana (ICCs range 89–.96; r scores range .81– 96). Test-retest reliability was also quite good for the total number of marijuana joints used (ICC = .78–.94; r scores range .79–95) and number of joints used per day (ICCs = .85–.93, r scores range .80–.94; Robinson et al., 2012)
- The TLFB has very good test-retest reliability for drinking, illicit drug use, and psychosocial functioning (r score > .90; Tonigan, Miller, & Brown, 1997). The TLFB shows good test-retest reliability over 5 days among substance-involved outpatients and for 30, 60, and 90 days across a range of drinking variables (Carey, Carey, Maisto, & Henson, 2004; Pedersen & LaBrie, 2006)

Concerns

- Completion time for the TLFB depends on the time period covered and the individual pattern of consumption
- There are lower agreement rates on the TLFB for shorter recall periods (e.g., shorter number of days assessed; Hjorthøj et al., 2012)
- The quantity of drug use may not be adequately assessed for drugs such as cocaine and amphetamines. A related concern to cannabis/marijuana is that the type of measurement used (e.g., number of joints) may not adequately assess the amount consumed

Availability and Cost

The TLFB instrument is available online at no charge from the Nova Southeastern University, Center for Psychological Studies at the following site: http://www.nova.edu/gsc/online_files.html

Calendars, instructions, and method manuals for alcohol, drugs, and nicotine can be downloaded at no cost. The Timeline Followback-User's Guide is available for $29.95 from the Centre for Addiction and Mental Health at the following site: http://www.camhx.ca/Publications/CAMH_Publications/timeline_followbk_usersgd.html

Recommendations for Assessment of Substance Use and Treatment Matching

Information in this section provides a critical review of treatment matching approaches and a description of specific instruments that can be used for assessing and matching offenders who have CODs to appropriate services. The assessment instruments described in this section vary considerably in the level of detail provided for mental disorders and CODs. This analysis is based on a review of research examining the reliability and validity of these approaches and instruments, the relative cost of instruments, ease of administration of instruments, and potential for application within the justice system. Although summaries of instruments are based on DSM-IV criteria, instrument recommendations are based on the potential for alignment with the DSM-5 criteria to allow for a more seamless transition to the newly implemented DSM-5 diagnostic classification system. Recommendations for assessment of substance use and treatment matching instruments include those that address criminogenic needs (i.e., "dynamic risk factors") as articulated by the RNR theoretical model. Recommendations for substance use and treatment matching instruments in the justice system include the following:

1. The TCU short forms (e.g., TCUDS V, TCU CEST, TCU TRMA, TCU HLTH). These

forms address key criminogenic needs and psychosocial factors related to treatment intake and matching, and can be tailored according to the specific resources and assessment needs of a particular justice program or setting.

(and/or)

2. The TCU Criminal Justice Comprehensive Intake (TCU CJ CI), which can be used in settings that do not currently utilize a standardized intake instrument. The TCU CJ CI intake can be combined with other short forms to provide a full assessment and to assist in treatment matching.

The TCU short forms each take approximately 5–10 minutes to administer and score and can be administered by nonclinicians who are trained in scoring and administration procedures and aware of appropriate referral procedures. The TCU CJ CI takes approximately 90 minutes to administer and score and should be conducted by a trained and licensed/certified clinician.

Assessment Instruments for Mental Disorders

The assessment instruments described below require significant training in administration, scoring, and interpretation. As a result, these instruments should be administered by trained mental health staff who are licensed, certified, or otherwise credentialed in assessing and diagnosing mental disorders and related psychosocial problems.

Minnesota Multiphasic Personality Inventory-2 (MMPI-2/MMPI-2 RF)

The MMPI (Hathaway & McKinley, 1951; Hathaway & McKinley, 1967; Hathaway & McKinley, 1989) is one of the most widely used instruments for assessment of mental disorders. The MMPI has been used in correctional settings since 1945 to classify individuals and to predict behaviors while incarcerated and after release (Megargee et al., 1979; Megargee & Carbonell, 1995). The MMPI-2 replaced the MMPI (Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989) following several rounds of scale revisions. The instrument is a self-report measure with 567 items and 10 main clinical scales, including Hypochondriasis, Depression, Hysteria, Psychopathic Deviancy, Masculinity-Femininity, Paranoia, Psychasthenia (obsessive-compulsive features), Schizophrenia, Hypomania, and Social Introversion. The MMPI provides 15 supplementary content scales that address internal traits, external traits, and general problems. In addition, the MMPI contains six validity scales that examine response sets, including unanswered items, endorsement of uncommon items, inconsistent responding, malingering, overreporting of symptoms, and faking good. An abbreviated version of the MMPI-2 includes 370 items, but scores obtained are not as comprehensive as the original 567-item version (Butcher & Hostetler, 1990). The MMPI-2 Restructured Clinical (RC) scales (Tellegen et al., 2003) are revised versions of the original clinical scales and improve upon the overlapping item content and high correlations between scales.

The most recent version of the instrument is the MMPI-2 Restructured Form (MMPI-2 RF; Ben-Porath & Tellegen, 2008), which is based on norms from the MMPI-2 and retains the same RC scales. The MMPI-2 RF has 338 items and 51 scales. These scales include Validity scales, Higher-Order scales (HO), RC scales, Somatic/Cognitive, Internalizing, Externalizing, Interpersonal, Interest, and Personality Psychopathology Five (PSY-5). Changes to the MMPI-2RF include improvement in the validity scales for nonresponding, inconsistent responding, overreporting, and underreporting of symptoms. The "?" or "cannot say" scale (CNS) has not been altered from the MMPI-2.

The MMPI-2 RF features revised versions of the MMPI-2 validity scales, including the following: Variable Response Inconsistency (VRIN-r) and

True Response Inconsistency (TRIN-r); the Lie scale, which is now Uncommon Virtues (L-r); and the K-Scale (Correction Scale), now referred to as Adjustment Validity (K-r). The latter two scales address underreporting of symptoms. The other four validity scales address overreporting of symptoms and improve upon three of the MMPI-2 scales of Infrequent Response (F-r), Infrequent Psychopathology Responses (Fp-r), and Symptom Validity (FBS-r, previously Fake Bad Scale; Ben-Porath, Tellegen, & Graham, 2008). An additional scale, the Infrequency Somatic Response (Fs) was added to identify overreporting of somatic complaints. The final scale, the Response Bias Scale (RBS; Gervais, Ben-Porath, Wygant, & Green, 2007), identifies overreporting in personal injury or medical disability settings and negative response bias in forensic settings.

All revised scales are shorter than the original validity scales and feature improved psychometric methods for testing the validity of these scales in detecting inconsistent responding and underreporting or overreporting of symptoms. The MMPI-RF T scores are not K-corrected (correction used to represent the accuracy of scores and to compensate for faking good or faking bad) nor are they gender specific. This allows for clinician judgment when examining differences between the non-K corrected clinical scale T scores and the K-corrected clinical scale T scores because previous research indicates that the K-corrected scales have poor validity. The RC (Restructured Clinical) scales are the same as those in the MMPI-2.

The MacAndrew Alcoholism Scale-Revised (MAC-R) was developed to differentiate alcoholic from nonalcoholic psychiatric patients. This supplementary scale on the MMPI-2 includes 49 items that provide a subtle screening measure to differentiate alcoholics from nonalcoholics (Searles et al., 1990). A 13-item Addiction Acknowledgment Scale (Weed, Butcher, McKenna, & Ben-Porath, 1992) was developed using items in the MMPI-2 whose content is clearly related to substance use. The Addiction Potential Scale was also developed, which included heterogeneous items related to extroversion, excitement seeking, risk taking, and lack of self-efficacy.

The MMPI-2 Criminal Justice and Correctional Report was developed for use in justice settings. This report assists in determining diagnoses and analyzing the MMPI-2 validity, clinical, content scales, and supplementary scales. The report provides information relevant to assessment, risk assessment, and treatment and program planning for individuals involved with the justice system. The report contains several behavioral dimensions that examine the need for further mental health assessment, conflict with authorities, extroversion, likelihood of favorable response to academic or vocational programming, and hostile peer relations. Several potential problem areas are also identified, related to alcohol or substance use, manipulation of others, hostility, and anger control.

Positive Features

- Only a sixth-grade reading level is required
- The MMPI-2 was normed using a large sample that was representative of the U.S. population
- A specialized interpretive report is available for justice-involved individuals
- Scales and profile configurations, which indicate personality profiles, have similar correlates in forensic settings as in other settings (Graham, 2006)
- The MMPI-2 has been used extensively with justice-involved individuals (Claes, Tavernier, Roose, Bijttebier, Smith, & Lillenfeld, 2012; Mattson, Powers, Halfaker, Akeson, & Ben-Porath, 2012; Wilson, 2012)
- The MMPI-2 is available in several languages and can be administered using a paper and pencil format, by audio recording, or via a computerized version of the instrument

- The MMPI-2 is well validated in a variety of settings and has good psychometric properties (Butcher, Graham, Ben-Porath, Tellegen, & Dahlstrom, 2001; Graham, 2000; Greene, 2000)
- A derived MMPI-RF measure of psychopathy corresponds well with other validated measures (e.g., Psychopathic Personality Inventory; Lilienfeld & Andrews, 1996) and traits (antisocial behaviors, narcissism; Phillips, Sellbom, Ben-Porath, & Patrick, 2014; Sellbom, Ben-Porath, Lilienfeld, Patrick & Graham, 2005; Sellbom et al., 2012)
- The MMPI-2 RC scales demonstrate concurrent validity with other similar substantive measures (Tellegen, Ben-Porath, & Sellbom, 2009). For example, RC2-low positive emotion is correlated with depressive mood symptoms (Arbisi, Sellbom, & Ben-Porath, 2008; Forbey & Ben-Porath, 2007; Handel & Archer, 2008) and social anxiety (Forbey & Ben-Porath, 2008), and RC1-somatic symptoms are correlated with somatoform problems (Arbisi et al., 2008; Forbey & Ben-Porath, 2007, 2008)
- The MMPI-2 RC scales indicate high internal consistency across gender groups in clinical representative samples (alphas range .78–.95; Rogers, Sewell, Harrison, & Jordan, 2006). The RC scales show improvement over the clinical scale in reducing interscale correlations (Rogers, Gillard, Berry, & Granacher, 2011; Tellegen et al., 2003)
- Several studies support the validity of the revised or added RF validity scales for the MMPI-2RF. The VRIN-r, TRIN-r, L-r, and K-r are useful in identifying underreporting among both clinical and nonclinical samples (Sellbom & Bagby, 2008). The Fp-R indicates incremental utility in detecting overreporting of psychopathology (Tellegen & Ben-Porath, 2008. The Fs scale also provides incremental utility in identifying exaggerated or "faked" somatic complaints (Wygant et al., 2007). The FBS-r, F-r, and F-s are able to identify neurocognitive, emotional, and somatic complaints (Wygant et al., 2010). Among offenders, the F-r and Fp-r were able to identify malingering of psychopathology (Sellbom, Toomey, Wygant, Kurcharski, & Duncan, 2010; Wygant et al., 2011), and these scales have been shown to be effective when compared to the Structured Interview of Reported Symptoms (SIRS; Rogers, Bagby, & Dickens, 1992)
- The Response Bias Scale (RBS; Gervais et al., 2007) is able to identify the validity of reported symptoms in forensic settings as demonstrated by its discriminatory ability to distinguish between those who pass or fail the symptom validity tests (Word Memory Test: Green, 2003; Test of Memory Malingering: Tombaugh, 1996). The RBS scale is also associated with other symptom validity scales such as the F-r, Fp-r, and Fs. Combinations of these scales can improve the specificity of overreported psychopathology and somatic complaints (Wygant et al., 2010)

Concerns

- The MMPI-2 requires somewhat more time to administer than the PAI
- The MMPI-2 RF does not include updated norms and is based on norms from the MMPI-2. Many validation studies of the MMPI-2RF employ the original validation data for the MMPI-2, and few studies have been conducted by those other than the instrument developers
- The MMPI-2 RC scales provide poor convergent validity for related areas of psychopathology (Rogers et al., 2011)
- Clinical elevations on the RC scales are difficult to interpret when used in combination, as scales can provide contradictory information. For example, RC1 demonstrates clinical elevations in over 60 percent of cases (somatic complaints), but these profiles were classified as within normal limits. The RCd, which reflects general psychiatric

distress, shows no elevation for those who endorsed persecutory ideation on RC6 (Rogers et al., 2011)

- Although the RBS scale improves identification of symptom validity, other symptom validity tests are still recommended during the assessment process (Heilbronner et al., 2009)
- The FBS-r and Fs may not perform well in detecting malingering, as they are focused more on somatic and cognitive deficit complaints (Sellbom et al., 2010)
- Many of the studies that validate scales of the MMPI-2 RF use archival data sets that have previously been used in validating the MMPI-2 and thus employ convenience sampling rather than replication in diverse samples
- Since the MMPI-2 is based on psychological constructs developed in the 1940s, both the content and clinical scales are somewhat heterogeneous. As such, there is some overlap among scales, which lessens the discriminant validity of this measure. For example, while it is possible to differentiate between bipolar disorder and schizophrenia using the Depression (Dep) content scale, no clinical or content scales on the MMPI-2 are able to differentiate between bipolar depression and unipolar depression (Bagby et al., 2005)
- The K correction scale does not have empirical support in many populations (Barthlow, Graham, Ben-Porath, Tellegen, & McNulty, 2002), and there is some disagreement regarding the cut-off scores to use for different validity scales to detect malingering (Meyers, Millis, & Volkert, 2002)
- Hispanic respondents produce higher scores on the Lie scale, and culturally specific norms or corrections have not been developed for this scale
- The MMPI-2 scale names do not reflect the domains that are measured
- The MMPI was developed using an empirical approach with the goal of discriminating between individuals with psychiatric diagnoses and individuals without any diagnosis. However, items were not selected based on theory or psychopathology research
- The MAC-R scale does not have good internal consistency (.56 for men and .45 for women; Butcher, Dahlstrom, Graham, Tellegen, & Kaemmer, 1989). In addition, several studies have urged caution when using the MAC-R scale with African Americans (Graham, 2006)

Availability and Cost

Information describing the MMPI-2 RF can be found at the following location, including scales, frequently asked questions, references, and an interpretation guide: http://www.upress.umn.edu/test-division/MMPI-2-RF/mmpi-2-rf-publications

The MMPI-2 RF manual, scoring sheets, and scoring/interpretive software can be purchased at the following location and are quite costly: http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAg523

Millon Clinical Multiaxial Inventory-III (MCMI-III)

The MCMI-III (Millon, 1983, 1997) is an objective, self-report psychological assessment measure consisting of 175 true/false items. The MCMI is designed to assess DSM-IV Axis II (personality) disorders and related clinical syndromes (Axis I) and is particularly useful in identifying personality disorders that may affect involvement in treatment. The Personality Inventory consists of 14 Personality Disorder Scales and 10 Clinical Syndrome Scales, both of which include separate Moderate and Severe Syndrome Scales. In addition, there are Correction Scales that help detect random responding and consist of three modifying indices (disclosure, desirability, and debasement) and one validity index. The MCMI-III contains three

Facet Scales for each MCMI-III Personality Scale. The Facet Scales were developed using factor analytic techniques and are included to guide clinicians in the interpretation of the Clinical Personality Patterns and the Severe Personality Pathology Scales. The scales aid in identifying specific personality processes (e.g., self-image, interpersonal conduct, cognitive style) that contribute to overall scale elevations. Base rates of disorders in the specific population are used as cut-off scores to indicate clinically significant levels of severity (i.e., > 75 percent = moderate level, > 85 percent = severe level; Millon, 1997).

Two of the Moderate Syndrome Scales of the MCMI-III address substance use (B-Alcohol Dependence, T-Drug Dependence). The MCMI-III is well suited for use in correctional settings. A separate Correctional Summary includes the use of special correctional norms for certain scales and a one-page summary of likely needs and behaviors relevant to corrections settings, including the need for mental health and substance use treatment. The report classifies a justice-involved individual's probable needs as low, medium, or high in the areas of mental health intervention, substance use treatment, and anger management services. In addition, escape risk, reaction to authority, disposition to malinger, and suicidal tendencies are evaluated.

Positive Features

- The MCMI-III is brief to administer, requiring approximately 25 minutes to complete
- The MCMI-III provides an interpretive report that describes potential DSM-IV diagnoses that may apply
- The instrument can be administered via paper and pencil, audiotape, CD, or computer
- The instrument is available in English and Spanish
- The measure was normed with adult inpatient and outpatient clinical samples and with individuals in jail and prison
- The MCMI-III has been used in justice/forensic settings (Bow, Flens, & Gould, 2010; Ferragut, Ortiz-Tallo, Loinaz, 2012; Morgan, Fisher, Duan, Mandracchia, & Murray, 2010; Young, Wells, & Gudjonsson, 2011)
- The AUC, sensitivity, and specificity are acceptable for the MCMI-III as determined by comparison with clinician-rated DSM-IV diagnoses (Millon, 1997)
- AUCs (> .70) for the MCMI-III scales are adequate for alcohol, drug, psychotic (MCMI-III delusions scale only), and major depressive disorders when compared to DSM-IV diagnoses (Hsu, 2002)
- The MCMI-III personality disorder scales show relatively good convergent validity with the MMPI scales for most disorders (Rossi, Hauben, Van den Brande, & Sloore, 2003)
- The MCMI-III demonstrates adequate diagnostic accuracy for Axis I disorders in international settings when compared with results from the Mini International Neuropsychiatric Interview (MINI; AUCs > .70), with the exception of psychotic disorders (Hesse, Guldager, & Holm Linneberg, 2012). This same study supports the convergent validity of MCMI-III scales with other measures, such as the Beck Anxiety Inventory and the MINI
- Another international study indicates acceptable sensitivity for the anxiety scale of the MCMI-III (73 percent), as identified by diagnoses obtained from the MINI (Saulsman, 2011)
- The sensitivity and specificity of MCMI-III Scales B (alcohol) and T (drug) are significantly improved from equivalent scales on the MCMI I and MCMI II (Craig, 1997)
- The MCMI-III disclosure, desirability, and debasement validity scales are effective in detecting malingering among traumatic brain injury patients (Aguerrevere, Greve, Bianchini, & Ord, 2011)

Concerns

- Little research has been conducted to examine the cultural sensitivity of the MCMI-III
- An eighth-grade reading level is required, which may be problematic in some justice settings
- AUCs for the MCMI-III anxiety and dysthymia scales are quite poor in detecting DSM-IV anxiety disorders or dysthymia (Hsu, 2002)
- An international study found poor agreement between the MCMI-III and the MINI in diagnosing treatment-seeking people with substance use disorders (Hesse et al., 2012)
- Another international study of a mental health treatment-seeking population indicated poor sensitivity for the MCMI-II in detecting anxiety disorders, dysthymia, and major depressive disorder and poor specificity for anxiety disorders and dysthymia, as indexed by the MINI clinical interview (Saulsman, 2011). The MCMI-III also did not adequately distinguish between anxiety disorders and depressive disorders
- Several studies examining the validity of the MCMI-III (Millon, 1994; Millon, 1997) indicate significant differences in diagnostic accuracy and raise methodological concerns (Hsu, 2002; Millon, 1994; Millon, 1997; Retzlaff 1996) related to the impact of varying levels of clinician skills and uneven interviewing procedures
- Some MCMI-III scales do not perform better than chance in detecting mental disorders and may not adequately discriminate between diagnoses (Hsu, 2002)
- The MCMI-III thought disorder scale (SS) may reflect general psychiatric distress, and it is correlated with measures such as the Beck Anxiety Inventory and Montgomery Asberg Depression Rating Scale (MADRS; Hesse et al., 2012)
- Based on the MCMI-III manual, approximately 13 percent of people who randomly respond on the instrument have invalid and noninterpretable results (Charter & Lopez, 2002). This study also indicates that too few items may be contained in the validity scale of the MCMI-III
- The MCMI-III may underreport personality disorders among justice-involved individuals (Retzlaff, Stoner, & Kleinsasser, 2002)
- In prior versions of the MCMI, the Drug Abuse Scale was found to have poor sensitivity (39 percent) but high specificity (88 percent) in identifying people with substance use disorders (Calsyn, Saxon, & Daisy, 1990)

Availability and Cost

The MCMI, manual, and hand-scoring guide can be purchased at the following site: http://psychcorp.pearsonassessments.com/HAIWEB/Cultures/en-us/Productdetail.htm?Pid=PAg505

Costs for the MCMI vary depending on the desired format. Scoring software is available that provides interpretive reports.

Personality Assessment Inventory (PAI)

The PAI is a self-administered objective test of personality and psychopathology developed to provide information related to treatment planning and evaluation. Although the instrument was introduced more recently than the MMPI and the MCMI, it has received considerable attention by clinicians and researchers because of its rigorous methodology. The development of the PAI was based on a construct-validation framework that emphasized a rational and quantitative method of scale development. A strong emphasis is placed on a theoretically informed approach to the development and selection of items (Morey, 1998). Key areas examined by the PAI include response styles, clinical syndromes, interpersonal

style, treatment complications, and subject's environment.

The PAI instrument includes 344 items and 22 nonoverlapping full scales, with 4 validity scales, 11 clinical scales, 4 treatment consideration scales, and 2 interpersonal scales. Validity scales include inconsistent responding (ICN), infrequency of endorsed response (INF), negative impression management (NIM), and positive impression management (PIM). Clinical scales include separate measures for alcohol problems (ALC), drug problems (DRG), somatic complaints (SOM), anxiety (ANX), anxiety-related disorders (ARD), depression (DEP), mania (MAN), paranoia (PAR), schizophrenia (SCZ), borderline personality disorder (BOR), and antisocial personality disorder (ANT). Treatment consideration scales include aggression (AGG), suicide ideation (SUI), stress (STR), nonsupport or lack of social support (NON), and treatment rejection (RxR). Interpersonal scales include dominance (DOM) and warmth (WRM). A T score ≥ 70 on the clinical scales, treatment scales, and interpersonal scales indicates clinically significant problems. There are 27 critical items that indicate acute problems (e.g., suicidal ideation) for which follow-up with the client should be provided. The PAI requires approximately 50 minutes to complete (Morey, 2007).

Positive Features

- The PAI was standardized on a sample that matched the 1995 census on gender, race, and age (Morey, 1998)
- PAI test items and scales were empirically derived and are based on clinical research and personality theory (Morey, 1991)
- A Spanish version of the PAI is available
- Additional software for justice settings is available that is geared towards assessment of risk, psychological needs, and rehabilitation
- Validity scales allow the clinician to detect whether items are left unanswered, answers are inconsistent, infrequent items are endorsed, and whether attempts are made to provide an overly negative or positive impression
- Information regarding symptom severity is provided, which helps in developing assessment and treatment recommendations
- The PAI includes 27 critical items, chosen based on their importance as indicators of potential crisis situations. These items facilitate follow-up probes to examine the need for crisis or other clinical services
- An interpretative profile is provided with each report to guide the clinician in developing treatment approaches
- The PAI is widely used in justice settings and substance use settings (Boccaccini, Murrie, Hawes, Simpler, & Johnson, 2010; Boccaccini, Rufino, Jackson, & Murrie, 2013; Magyar et al., 2012; Patry, Magaletta, Diamond, & Weinman, 2011; Ruiz et al., 2012; Salekin, 2008; Walters, Duncan, & Geyer, 2003)
- The PAI is used in the criminal sentencing process, including cases involving capital sentencing (Mullen & Edens, 2008)
- The PAI-ANT scale is related to other measures of antisocial behaviors and criminal thinking (Bradley et al., 2007; Douglas et al., 2007; Walters & Geyer, 2005), such as the Shedler-Westen Assessment Procedure (SWAP-200; Westen & Shedler, 1999a,1999b), and measures of psychopathy (Douglas, Guy, Edens, Boer, & Hamilton, 2007; Patrick, Poythress, Edens, Lilienfeld, & Benning, 2006; Edens & Ruiz, 2005), such as the Psychopathy Checklist-Revised (PCL-R; Hare & Vertommen, 2003) and the Psychopathic Personality Inventory (PPI; Lilienfeld & Andrews, 1996)
- The ANT scale contains subscales examining aggression, dominance, and violence potential and provides an assessment of risk factors that predict recidivism and violence in offenders (Boccaccini et al., 2010; Morey, Warner, & Hopwood, 2007)

- The ANT, AGG, and DRG scales have been found to predict prison infractions in an international offender sample, including violent, nonviolent, and drug-related infractions and recidivism (Newberry & Shuker, 2012), as indexed by the Offender Group Reconviction Scale (OGRS, Copas & Marshall, 1998)
- Incremental validity for the PAI-ANT scale has been found in predicting disciplinary problems, verbal and physical aggression, and recidivism (Buffington-Vollum, Edens, Johnson, & Johnson, 2002; Walters & Duncan, 2005; Walters et al., 2003) in comparison to clinical measures such as the PCL-R (Hare & Vertommen, 2003). The scale performs as well as the Static-99 (Hanson & Thornton, 1999) and Minnesota Sex Offender Screening Tool-Revised (Epperson, Kaul, Hesselton, 1998) in predicting recidivism among sexual offenders (Boccaccini et al., 2010)
- In an offender sample, incremental validity has been found for the AGG scale in predicting noncompliance (e.g., gambling, stealing) and aggressive behaviors (both verbal and physical) above and beyond scales such as ANT and BOR. Overall, AGG, BOR, and ANT scales have been found to predict aggressive or disruptive behaviors (Magyar et al., 2012)
- The concurrent validity of the PAI with offenders is supported by findings indicating that the DRG and ALC scales are correlated with other indices of alcohol use and drug use from the Federal Bureau of Prisons mental health data base, psychological intake questionnaire, and presentencing reports (Patry et al., 2011)
- In support of the PAI's external validity among offenders who are court mandated to substance use treatment, higher scores on the AGG scale are correlated with a history of assault. Similarly, higher ANT scale scores are related to rule-breaking while in treatment, particularly among offenders who have higher scores on the DRG scale. The SUI scale accurately identifies those who have a history of suicide attempts (Hopwood, Baker, & Morey, 2008)
- Also supporting external validity of the PAI with both psychiatric inpatients and outpatients, the PAI clinical scales show moderate to strong correlations with life events that are relevant to PAI scales. For example, the ANT scale is correlated with history of arrest, alcohol, and drug problems, and lower education level. Similarly, the DRG, ALC, BOR, and AGG scales are correlated with the history of arrest. The ARD scale is also correlated to trauma and prior history of hospitalization, and the DEP scale is correlated with prior hospitalization (Slavin-Mulford et al., 2012)
- Within offender samples, the PAI clinical scales may reflect a two-dimensional structure of "internalizing" and "externalizing" tendencies, as indicated by statistical taxometric procedures and confirmatory factor analysis (Ruiz & Edens, 2008)
- The overall psychometric properties of the PAI are quite favorable (Morey, 1991; Morey, 2007) and include high internal consistency of scales (Magyar et al., 2012)
- Full-scale reliability estimates for the PAI are high, averaging .82 (Boone, 1998)

Concerns

- The PAI is a commercially available instrument
- Only trained mental health professionals can administer and interpret the PAI
- The PAI may be lengthy to administer, typically requiring an hour but sometimes requiring up to 2.5 hours to complete
- The Spanish version of the PAI may not provide psychometric properties that are equivalent to the English version (Fernandez, Boccaccini, & Noland, 2008; Rogers, Flores, Ustad, & Sewell, 1995)
- Several unique issues should be considered in interpreting the PAI's validity scales in justice and treatment settings. For

example, people seeking treatment may have higher NIM scale scores as they may exaggerate symptoms to secure treatment. PIM scores may also be elevated in justice settings as a result of attempts to deny potential problems, such as substance use (Douglas et al., 2007; Morey & Quigley, 2002; Newberry & Shuker, 2012). INF and ICN scores may also be inflated among offenders, who tend to respond inconsistently and to endorse items with low base rates (Douglas et al., 2007; Newberry & Shuker, 2012). However, scale scores may be affected by poor reading abilities (Nikolova, Hendry, Douglas, Edens, & Lilienfeld, 2012)

- Inappropriate use of cut-off scores with offenders may lead to misclassification in determining "risk" level and in assignment to services (Edens, Poythress, & Watkins-Clay, 2007)
- For offenders with high PIM scale scores (T scores ≥ 57), the violence potential index (composed of items from different PAI scales, including drug use, aggression, and antisocial behaviors) and the SUI and STR scales may not be useful in assessing risk, and ANT scale scores may not as effectively predict problem behaviors (Walters, 2007)
- The PAI's alcohol and drug scales are susceptible to denial since the item content is not subtle

Availability and Cost

The PAI is available at cost from Psychological Assessment Resources at the following site: http://www4.parinc.com/Search.aspx?q=PAI

There are numerous PAI resources available, including the instrument, scoring sheets, an interpretive guide, a user manual, and scoring software that generates interpretive reports. Supplementary software is also available that generates interpretive reports geared for correctional settings.

A PAI kit can be purchased for $315 and includes the professional manual, answer booklets, the instrument, and materials for hand scoring (e.g., profile forms).

Recommendations for Assessment of Mental Disorders

Information describing assessment instruments for mental disorders is based on a critical evaluation of the research examining the efficacy of these instruments. Important indicators used in evaluating instruments include the following: empirical evidence supporting both the reliability and validity of the instrument, ability to assess multiple mental health problems/disorders, the relative cost of the instrument, ease of administration and interpretation, and previous use within justice settings. Although the assessment instruments provide information that addresses the range of mental disorders described in the DSM-IV, it is highly desirable for these instruments to be closely aligned with the newly implemented DSM-5 criteria to allow for a seamless transition from the DSM-IV to DSM-5 diagnostic classification systems. Based on these considerations, the following instrument is recommended for use in assessing mental disorders for people with co-occurring disorders in the justice system:

- The Personality Assessment Inventory (PAI)

The PAI assesses personality traits, mental health problems/disorders, and other treatment-related problems and requires approximately 45–60 minutes to administer and 25–30 minutes to score and interpret. The PAI provides several validity indices and facilitates clinician follow-up to individual item responses. The PAI should be administered and interpreted by a trained and licensed/certified mental health professional.

Assessment and Diagnostic Instruments for Co-occurring Mental and Substance Use Disorders

This section reviews instruments that are used to diagnose or assess CODs. Included are assessment instruments that examine other biopsychosocial domains related to CODs. Diagnostic instruments include those that evaluate DSM or ICD disorders and provide a diagnosis for a range of mental and substance use disorders. Some instruments, such as the GAIN and MINI, which include multiple versions (e.g., screening, assessment) are described in this and other sections. In contrast to instruments described in screening sections, assessment instruments described in this section require more time to administer; provide more detailed and comprehensive coverage of issues related to the various disorders; and are designed to yield formal diagnoses and treatment plan recommendations, including levels and types of services that are needed. The assessment and diagnostic instruments described below require significant training in administration, scoring and interpretation. As a result, these instruments should be administered by trained clinicians who are licensed, certified, or otherwise credentialed in assessing and diagnosing mental and substance use disorders and related psychosocial problems.

Assessment Instruments for Co-occurring Mental and Substance Use Disorders

Alcohol Use Disorders and Associated Disabilities Interview (AUDADIS-IV)

The AUDADIS-IV (Grant & Dawson, 2000) is both an assessment and diagnostic instrument, and is a fully structured clinical interview that is based on the DSM-IV and ICD-10 criteria. The AUDADIS-IV assesses alcohol, drug, and nicotine use disorders. It also assesses mental disorders, including mood disorders, anxiety disorders, and DSM-IV personality disorders, in addition to the family history of mental disorders. The instrument is standardized to diminish the unreliability that is often found in other structured interviews and navigates complex diagnostic criteria by use of multiple short questions. If the respondent meets criteria for a particular diagnosis, all questions in the module are asked to provide a more complete dimensional assessment of related problems. The instrument requires approximately 1 hour to administer and provides both lifetime (prior to past 12 months) and current diagnoses (past 12 months). The AUDADIS-IV examines the onset of disorders; duration of symptoms of each disorder; the presence of co-occurring disorders; severity and impairment of symptoms, including "rule out" causes of symptoms (e.g., use of medication or drugs); frequency of substance use, patterns of use; and quantity of use. The most recent version of the AUDADIS-IV includes additional risk factor scales related to social and occupational functioning, such as the self-reported discrimination scales (e.g., reported bias against race, weight, ethnicity, culture). The instrument also examines stressful life events and perceived stress.

Positive Features

- The AUDADIS-IV is fully structured and translates DSM-IV criteria into simpler language and thus can be administered by nonclinicians
- The AUDADIS-IV has been translated into Spanish
- The AUDADIS-IV was designed to comprehensively assess for CODs among people who have substance use disorders
- The AUDADIS-IV provides adequate coverage of quantity, frequency, and duration of substance use disorders
- The AUDADIS-IV provides improved coverage of the chronology of symptoms and disorders in comparison to other structured assessment interview instruments (Grant et al., 2003)
- The AUDADIS-IV has been used with offenders to study antisocial behaviors and their correlates (e.g. drug use, low

income,) in a large national epidemiological survey (Gelhorn, Sakai, Kato Price, & Crowley, 2007; Hoertel, Le Strat, Schuster, & Limosin, 2012; Vaughn et al., 2011; Vaughn et al., 2010)

- The AUDADIS-IV has also been used as a diagnostic/assessment tool in justice settings (Kerridge, 2009)
- The concurrent validity of the AUDADIS-IV is supported by findings of high comorbidity of nicotine disorders with other substance use disorders and is correlated with mental health scores on the SF-12; (Short Form Health Survey, Compton, Thomas, Stinson, & Grant, 2007; Gandek et al., 1998; Grant et al., 2004; Hasin, Stinson, Ogburn, & Grant, 2007; Kessler et al., 1994)
- The concurrent validity of the AUDADIS-IV is also supported by findings from a large epidemiological study that yielded high rates of co-occurring substance use, anxiety, and mood disorders (Grant et al., 2004). This same study indicated that personality disorders were associated with lower mental health scores as measured by the SF-12 (Grant et al., 2004). Borderline personality disorder was associated with increased mental and social difficulties, which is consistent with findings from other studies (Grant et al., 2008)
- Concurrent validity is also supported by findings of high rates of co-occurring depression among offenders who have substance use disorders (Kerridge, 2009)
- In large representative samples, interrater reliability for drinking and tobacco use frequency and quantity were quite good over an average 10-week period, with ICCs ranging .69–.84 (Grant, Dawson, Stinson, Chou, Kay, & Pickering, 2003). Interrater reliability for current and lifetime alcohol use disorders is also quite good (kappas range .70–.74; Grant et al., 2003)
- Interrater reliability for depressive disorders is acceptable (kappas range .59–.65), and reliability for severe anxiety is quite good (ICCs range .71–.86). Interrater reliability for adult ADHD and current/lifetime PTSD is adequate (kappas range .63–.77; Ruan et al., 2008)
- The Spanish version of the AUDADIS-IV demonstrates good psychometric properties, including test-retest reliability and interrater reliability for agreement on diagnoses (Mestre, Rossi, & Torrens, 2013)
- Internal consistency of the additional risk factor scales related to perceived stress and stressful life events are good (alphas range .82–.94), and discrimination for current/lifetime symptoms is acceptable (alphas range .59–.78; Ruan et al., 2008)

Concerns

- The AUDADIS-IV was developed in the general population and would benefit from further validation in clinical, criminal justice, and substance use settings
- Further validation is needed for AUDADIS-IV modules examining PTSD and DSM-IV personality disorders
- The AUDADIS-IV does not assess for psychosis other than inquiring about lifetime diagnosis of schizophrenia and assessment of schizoid personality disorder (Grant et al., 2003)
- The AUDADIS-IV may not effectively diagnose current/lifetime anxiety disorders (ICCs range .40–.52, Grant et al., 2003)
- The discrimination scales indicate relatively low internal reliability across current and lifetime time periods (Ruan et al., 2008)

Availability and Cost

The AUDADIS-IV is available free of charge and can be obtained by contacting Dr. Bridget Grant at bgrant@willco.niaaa.nih.gov

The Composite International Diagnostic Interview (CIDI)

The CIDI is a structured comprehensive interview developed by WHO to assess mental disorders

according to the definitions and criteria of the International Classification of Disease (ICD, ICD-10) and the DSM (DSM-IV). The CIDI is one of the most widely used structured diagnostic interviews internationally, as it was developed specifically for use among different cultures and settings. The instrument was derived from the Diagnostic Interview Schedule (DIS; Robins, Helzer, Croughan, & Ratcliff, 1981) and accommodates diagnoses based on the definitions and criteria of both the ICD and DSM. The CIDI was first used in 1990 and was revised and expanded in 1998 by the WHO World Mental Health (WMH) initiative to address subthreshold impairment, symptom severity and persistence, risk factors, internal and external (global) impairment, consequences, patterns of treatment, and treatment adequacy, in addition to diagnosis of mental disorders (Kessler & Üstün, 2004). The WMH-CIDI contains 22 diagnostic sections, including anxiety, mood, eating, tobacco, and substance use disorders, attention deficit hyperactivity disorder (ADHD), conduct disorder, psychosis, and personality disorders. There are four sections assessing functioning and physical comorbidity, two sections assessing treatment, seven sections assessing sociodemographics, and two sections assessing methodological factors (e.g., interviewer observations). The CIDI-SAM (Substance Abuse Module) can be used separately, if desired, to diagnose substance use disorders.

Positive Features

- Administration of the CIDI does not require use of mental health professionals or significant clinical training to administer
- The CIDI provides both ICD-10 and DSM-IV diagnoses
- A diverse sample was used to develop the instrument, including individuals with a broad range of alcohol and drug use severity
- The WMH-CIDI has been translated into several languages using the standard WHO translation and back-translation protocol
- A computerized version of the CIDI is available, which contains a scoring algorithm to provide a diagnosis. The computerized version has the ability to handle more elaborate "skip" patterns, while covering the same information as the paper and pencil version (WHO, 2004)
- The CIDI has been used to diagnose disorders among people with intoxicated driving charges (Lapham, Baca, McMillan, & Lapidus, 2006; Shaffer et al., 2007), prisoners (Brinded, Simpson, Laidlaw, Fairley, & Malcolm, 2001), and juvenile offenders (Steinberg, Blatt-Eisengart, & Cauffman, 2006)
- The CIDI-SAM shows acceptable agreement with the Schedules for Clinical Assessment in Neuropsychiatry (SCAN; Wing et al., 1990) in diagnosing alcohol use disorders (kappa = .69) and cocaine use disorders (.61; Compton, Cottler, Dorsey, Spitznagel, & Mager, 1996). A nationally representative U.S. survey also indicates positive findings for the AUC for the WMH-CIDI for substance use disorders (AUC = .72–.99), anxiety disorders (AUC = .74–93), mood disorders (AUC = .87–.97), and "any" disorder (AUC = .76; Haro et al., 2006). According to this same survey, the CIDI-SAM demonstrates good test-retest reliability for substance use disorders over a 1-week period (kappas range 63–.80; Horton, Compton, & Cottler, 2000)
- The CIDI has good sensitivity (74 percent) and specificity (98 percent) for any substance use diagnosis (Haro et al., 2006) and has adequate sensitivity for anxiety disorders (84 percent), mood disorders (69 percent), or "any" disorder (78 percent). The CIDI has excellent specificity (93 percent, 97 percent, and 91 percent for each of these respective disorders; Haro et al., 2006), and good positive predictive values and negative predictive values
- The WMH-CIDI demonstrates good sensitivity, specificity, positive predictive values, and negative predictive values

across different mental disorders and severe substance use disorders (Kessler et al., 1998), although the reliability of substance use diagnoses have been less than adequate in several studies (Kessler et al., 1998; Üstün et al., 1997)

- The WMH-CIDI provides adequate agreement with the SCID-I for substance use diagnoses (Haro et al., 2006)

Concerns

- The CIDI is quite lengthy and requires an average of 2 hours to administer
- Use of the WMH-CIDI requires completion of a training program that reviews interviewing techniques and field quality control
- In a large U.S. survey, the WMH-CIDI exhibited low accuracy in identifying substance use disorders and a range of mental disorders when compared with the SCID-I (Haro et al., 2006)
- Little data is available regarding the CIDI's effectiveness in justice settings

Availability and Cost

Both printable to paper and computerized versions of the CIDI can be obtained free of charge from the World Health Organization at the following site: http://www.hcp.med.harvard.edu/wmhcidi/instruments_download.php

Information regarding training in use of the CIDI can be found at the following site: http://www.hcp.med.harvard.edu/wmhcidi/trc_main.php

Global Appraisal of Needs (GAIN)

The 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. A more detailed description of the GAIN family of instruments is provided in the section entitled, "Screening Instruments for Co-occurring Mental and Substance Use Disorders." The GAIN instruments can be administered via interview or self-administered by paper and pencil or by computer. A wide variety of software is available to score and interpret results of the GAIN instruments. The Quick version of the GAIN (GAIN-Q3) requires 25–35 minutes to administer and includes assessment of nine individual sections related to a wide range of psychosocial and behavioral health issues in adults and adolescents. The GAIN examines areas such as substance use, mental health status, physical health, stress, work problems, life satisfaction, behavioral problems, and service utilization in the past 90 days. The GAIN instrument can also be used as a follow-up tool to assess and monitor progress. The GAIN-Q provides a recommended cut-off score of ≥ 3 for both adults and adolescents in identifying people with a mental disorder (Dennis et al., 2006). Other versions of the instrument include the GAIN-Q3-Lite, which consists of nine individual screeners and requires approximately 25 minutes to administer. The GAIN-Q3-MI (motivational interviewing) includes information regarding readiness for treatment and change.

The GAIN-Initial requires approximately 120 minutes to administer and provides a full assessment of psychosocial issues related to substance use treatment, as well as internalizing and externalizing disorders and problems related to crime and violence. The GAIN-Initial is useful for diagnostic purposes, treatment planning, placement in different levels of treatment services, and monitoring offender and/or program outcomes. Several versions of the GAIN-Initial have been developed for various programs, primarily those funded by CSAT and by the Robert Wood Johnson Foundation. Several follow-up forms are available to examine change over time in psychosocial areas related to treatment. The GAIN-I Lite is shorter to administer, requiring approximately 60 minutes, but is not as detailed as the full version. It contains the GAIN-Q3, other items needed for diagnosis, and the American Society of Addiction Medicine (ASAM) placement criteria for treatment planning and referral. The GAIN-I Core is used when the GAIN-Initial

cannot be administered and contains less detailed information examining service utilization and treatment history. The GAIN-I core requires 60–75 minutes to administer. The GAIN-M90 monitors treatment progress and is administered at 6, 9, and 12 months following treatment initiation; it requires approximately 60 minutes to administer.

Positive Features

- The GAIN-Q and GAIN-I is designed for use in justice settings, primary care settings, substance use treatment programs, and other social service programs
- Norms for the GAIN have been developed for adults and adolescents and for different levels of care. Additional norms are being developed by gender, race/ethnicity, CODs, and for juvenile and adult offenders
- Scoring software is available to interpret scores for purposes of diagnosis and treatment planning. Personal feedback reports (PFR) are also available
- Computerized versions of the GAIN are available that provide interpretation of assessment and validity reports to identify erroneous or missing data. A wide variety of support services are available through the GAIN Coordinating Center
- The GAIN has been used to assess mental disorders among juvenile and adult offenders (Belenko, 2006; Hussey, Drinkard, & Flannery, 2007; Sacks et al 2007b, Ramchand, Morral, & Becker, 2009)
- The GAIN has been widely used to assess mental health problems among adolescents and adults enrolled in substance use treatment (Chan, Dennis, & Funk, 2008; Dennis, White, & Ives, 2009; Shinn et al., 2007)
- Among adults, the GAIN-I demonstrates good predictive utility related to recidivism and relapse (Dennis, Scott, & Funk, 2003; Dennis et al., 2006)
- The GAIN-I–Substance Problem Scale is correlated with increased risk of internalizing and externalizing disorders among adults. The Behavior Complexity Scale is correlated with severity of substance use problems, and the Crime/Violence Scale is correlated with future criminal behavior (Dennis et al., 2006)
- A confirmatory factor analysis supports the factor structure of the GAIN in adults, including its use as a unidimensional measure (total score) and use of the individual subscales (Dennis et al., 2006)
- The GAIN-I and its subscales have good internal consistency for use with adults, with alphas ranging .71–.96 (Dennis et al., 2006). Studies examining concurrent validity have been conducted primarily with adolescents, but are quite promising (Dennis et al., 2006)
- The GAIN-Q and its subscales have adequate internal consistency among adults (GAIN Coordinating Center, 2012)
- The GAIN-I demonstrates good internal consistency for three comorbidity subscales related to internal mental distress, behavior complexity, and crime/violence, with alphas ranging .78–.96. The condensed versions of these scales, the internal behavior scale, and the external behavior scale also demonstrate good internal consistency, with alphas ranging .69–.90 (Titus, Dennis, Lennox, & Scott, 2008). The GAIN original scales are highly correlated with the subscales for adults
- The GAIN-I has good test-retest reliability for the main subscales (internal mental distress, behavior complexity scale, substance problem scale, crime/violence scale), with r score = .70 and kappas = .60. The GAIN-I also has good agreement with timeline followback, urinalysis, treatment, and other measures of substance use disorders (r score $\geq$ .70 and kappa $\geq$ .60; Dennis et al., 2006)
- Among adolescents, the GAIN-I shows good agreement with diagnoses of ADHD, mood disorders, conduct disorder/oppositional defiant disorder, and adjustment disorder and distinguishes

between co-occurring psychopathology (kappas range .65–1.00; Shane, Jasiukaitis & Green, 2003)

- Among adolescents, the GAIN-I has good internal consistency for three subscales of internal mental distress, behavior complexity, and crime/violence (Dennis et al., 2006; Titus et al, 2008). Original scales were highly correlated with shortened subscales among both adults and adolescents (Titus et al., 2008)

Concerns

- Training is strongly recommended before administering the GAIN. The GAIN training is costly and includes separate trainings to administer the instrument and to train others on how to use the measure
- The GAIN is a copyrighted instrument, and there are separate costs to purchase the set of instruments and for the software
- License agreement paperwork and a separate user agreement are required at cost
- Further validation among offender populations is needed to examine the GAIN's psychometric properties, including predictive utility of diagnoses and diagnostic impressions. Self-reported substance use on the GAIN is only moderately correlated with drug testing and other collateral information (Dennis et al., 2006)
- Item response theory (IRT) analyses show that the crime/violence scale on the GAIN may be less reliable for adults, particularly among adult females, potentially leading to errors in clinical diagnoses (Conrad et al., 2010)

Availability and Cost

A license agreement and separate user agreement is required ($100), which provides up to 5 years of coverage and unlimited paper assessments. License agreements can be ordered by e-mail at gaininfo@chestnut.org or by calling (309)-451-7900.

Scoring and diagnostic interpretation using the paper version of the GAIN-I and GAIN-Q are described in the GAIN manual. Using the hand-scored approach requires substantially more time than automated scoring provided using the web version. The various GAIN manuals can be obtained at the following locations:

GAIN-I: http://gaincc.org/_data/files/Instruments percent20and percent20Reports/Instruments percent20Manuals/GAIN-I percent20manual_ combined_0512.pdf

GAIN-Q: http://www.gaincc.org/_data/files/ Instruments percent20and percent20Reports/ Instruments percent20Manuals/GAIN-Q3_3.1_ Manual.pdf

The GAIN-ABS (Assessment Building System) is an online system that provides administration, scoring, and interpretative reports for the GAIN-I and GAIN-Q3. This version requires the license agreement as noted above, in addition to separate user agreements. The cost is $180 per user/ per year in addition to a one-time $100 start-up fee. There are standardized costs established for groups of users. A 30-day free trial period is also available. Interpretative reports are only available using the web version of the GAIN. Administration training costs range from $1,200 to $1,800. Different training is provided to administer the GAIN-I and GAIN-Q3. Training recipients are not authorized to train others on how to administer the instrument. Local training certification is provided for those who would like to train other users. These certificates cost between $1,500 and $2,400 for the GAIN-I and GAIN-Q3. Each type of training is available online; however, there are designated time limits in which the training must be completed (i.e., 3–6 months).

Information describing the GAIN-Q administration, scoring, and norms can be found at the following site: http://www.gaincc.org/index.cfm?pageID=51

Information describing the GAIN-I administration, scoring, and norms can be found at the following site: http://www.gaincc.org/products-services/instruments-reports/gaini/

Diagnostic Instruments for Co-occurring Mental Health and Substance Use Disorders

Diagnostic Interview Schedule–Fourth Edition (DIS-IV)

The DIS-IV is a fully structured diagnostic interview instrument designed for research purposes (Blouin, Perez, & Blouin, 1988; Robins et al., 1981) and has been updated to coincide with revisions to diagnostic categories in the DSM. Revised versions of the DIS have improved accuracy in identifying a range of mental disorders. A self-administered computerized version of the DIS is available (C-DIS), although staff must be present to address respondents' questions. Administration of the DIS does not require clinical experience. The DIS-IV has 19 diagnostic modules covering over 30 Axis I disorders, which include demographic and risk factors, sequencing of comorbid disorders, observations of psychotic symptoms or other problems during the interview, and a range of individual modules examining different types of disorders related to mood, anxiety, eating, schizophrenia spectrum, somatization, substance use disorders, antisocial personality disorder, ADHD, dementia, and gambling. The DIS provides information regarding both current and lifetime diagnoses of common mental disorders.

Positive Features

- The DIS can be administered by nonclinicians, requires minimal training, and has been translated into many languages
- The DIS has been used to diagnose mental disorders among offenders (Lo & Stephens, 2000; Teplin et al.,1996; Wiesner, Kim, & Capaldi, 2005) and people with substance use disorders (Havassy, Alvidrez, & Owen, 2004; Horton, Compton, & Cottler, 1998)
- In addition to detecting the presence of mental disorders in the justice system, the DIS has been used to refer offenders to treatment (Lo, 2004; Teplin, 1990)
- The DIS includes an antisocial personality disorder (ASPD) module. DIS-IV diagnoses of ASPD are correlated with substance use and chronic patterns of offending (Wiesner et al., 2005)
- The DIS has good agreement with the MAST (.79) in detecting alcohol disorders among individuals treated for mental disorders (Goethe & Fisher, 1995). Reliability of DIS diagnoses is quite good because interview questions, probes, and coding procedures are carefully described (Compton & Cottler, 2004)
- The DIS has adequate agreement with the SCAN for diagnosis of substance use disorders and for depression (Compton & Cottler, 2004) and has excellent specificity (90 percent) in detecting depression (Eaton Neufeld, Chen, & Cai, 2000)
- The DIS demonstrates adequate agreement with medical chart diagnoses (Robins, Helzer, Ratcliff, & Seyfried, 1982)
- The DIS diagnoses provide adequate agreement with most lifetime disorders, as determined by the DSM-III-R among psychiatric patients (kappas ≥ .5; Robins et al., & Ratcliff, 1981; Robins et al., 1982). Similarly, in college students, interrater agreement for both current and lifetime disorders on the DIS is acceptable (median kappas range .43–.46; Vandiver & Sher, 1991)
- Wittchen et al. (1989) found good agreement (kappas range .50–.70) between the clinician-administered and nonclinician-administered interviews for the DIS, as well as good test-retest reliability between administrations of the DIS (kappa > .6).
- The DIS has good test-retest reliability (95 percent agreement for severe disorders) in

diagnosing men who are incarcerated in jail (Abram & Teplin, 1991)

Concerns

- The DIS is quite lengthy, requiring 90–120 minutes to administer. However, it is possible to omit sections of the DIS that are not of interest
- Further validation of DIS diagnoses is needed with offenders
- Structured instruments such as the DIS may fail to detect 25 percent of those abusing alcohol (Drake et al., 1990) and possibly a higher proportion who are abusing illicit substances (Stone, Greenstein, Gamble, & McLellan, 1993)
- There is poor agreement between the DIS and the Schedule for Affective Disorders and Schizophrenia- Lifetime (SADS-L) in diagnosing depression among individuals who have CODs (Hasin & Grant, 1987)
- The DIS may be overly sensitive in diagnosing major depressive disorder (Helzer et al., 1985)
- The DIS has low agreement with the SCAN for diagnosis of depression (Eaton et al., 2000)
- The DIS may not accurately diagnose anxiety disorders (e.g., panic, social phobia) or schizophrenia (Anthony et al., 1985; Cooney, Kadden, & Litt, 1990; Erdman et al., 1987; Summerfeldt & Antony, 2002)
- Caution is urged when using the DIS as a primary diagnostic tool, as agreement between the DIS and clinician diagnosis has sometimes been poor in comparison to that of the SCID (Blanchard & Brown, 1998)
- The C-DIS provides poor to moderately good (-.05–.70) test-retest reliability in diagnosing CODs, depending on the type of mental disorder (Ross, Swinson, Doumani, & Larkin, 1995)
- The DIS is not sensitive to response styles and does not provide methods for detecting dissimulation (Alterman et al., 1996)

Availability and Cost

A copy and license for the use of the DIS (computerized version) may be purchased at the following site: http://epidemiology.phhp.ufl.edu/assessments/c-dis-iv/brochure/

The cost for licensing ranges from $1,000 to $2,000.

The Mini International Neuropsychiatric Interview (MINI)

The MINI (Sheehan et al., 1998) is a 120-question structured diagnostic interview used to evaluate DSM and ICD Axis I mental disorders (although the DSM-5 does not have axes, some of these frameworks are built around DSM-IV and earlier versions), including substance use disorders. The instrument was designed as a brief diagnostic screen and has been used in numerous research and clinical settings. The MINI provides a family of structured interviews, which includes the MINI, MINI-Kid, MINI-Plus, and MINI-Screen. Another section, "Screening Instruments for Co-occurring Mental and Substance Use Disorders," provides a more detailed description of the MINI screening tool. The MINI-Plus is a fully structured instrument that assesses the presence of 23 DSM-IV-TR Axis I disorders, including attention deficit hyperactivity disorder (ADHD) and one Axis II disorder (antisocial personality disorder), chronology of disorders, and rule-out questions to accurately identify the presence of comorbid disorders. The Mini-Kid screens for common childhood and adolescent psychopathology, including mood disorders, anxiety disorders, substance use disorders, externalizing disorders, and developmental disorders. Other MINI instruments have been developed to examine bipolar and psychotic disorders and suicidality. The most recent version of the MINI, MINI 6.0, is also available for administration by computer.

Positive Features

- Only brief training is required to use the instrument
- The MINI provides a diagnostic impression for major "Axis I disorders" and examines a broad range of symptoms. The instrument requires approximately 20 minutes to administer to individuals who do not have a mental disorder
- The MINI has been translated into many languages and includes norms for several subpopulations (Sheehan et al., 1998)
- The MINI-Plus has been used with offenders to assess current and lifetime mental and substance use disorders (Black et al., 2007; Cuomo, Sarchiapone, Di Giannantonio, Mancini, & Roy, 2008; Gunter et al., 2008), including antisocial personality disorder (Black, Gunter, Loveless, Allen, & Sieleni, 2010). In a study of the MINI-Plus with a prison sample (Black et al., 2004), the measure was easily administered by correctional staff, well received by prisoners, and it accurately assessed mental disorders in this population
- The MINI clinician-administered interview demonstrates good sensitivity (62–96 percent) and specificity (86–100 percent) across almost all current/lifetime Axis I disorders as determined by the SCID-I patient clinical interview (Sheehan et al., 1998). Similarly, the MINI patient rated self-report instrument has adequate sensitivity (60–89 percent) and good specificity (74–99 percent) for many of the current/lifetime Axis-I diagnoses. The MINI also has good sensitivity (67–89 percent) and specificity (72–97 percent) for many CIDI (Composite International Diagnostic Interview) DSM-III-R disorders. Overall specificity is good for the MINI as compared to other structured clinical interviews (Sheehan et al, 1998)
- Agreement between MINI clinician-rated and CIDI diagnoses for psychotic disorders is adequate (kappas range .68–.82), as are those between the MINI and SCID–I diagnoses (Sheehan et al., 1998)
- Interrater reliability estimates for the clinician-administered version of the MINI ranges .79–1.00 for all subscales. Fourteen of the 23 test-retest reliability values are greater than .75 (range = .35–1.00, and only one is below .50; Sheehan et al., 1998)
- The MINI shows good concordance with SCID DSM-IV diagnoses (kappas range .90–1.0; Sheehan et al., 1998)
- The MINI-Kid shows good sensitivity (71–100 percent) and specificity (74–99 percent) in identifying mental disorders as determined by the K-SADS-PL (Schedule for Affective Disorders and Schizophrenia for School-Aged Children; Kaufman et al., 1997). For individual diagnosis, sensitivity is adequate (67–100 percent) and specificity (73–99 percent) is good across most disorders (Sheehan et al., 2010). Interrater reliability for the MINI-Kid is also good (Sheehan et al., 2010)
- Test-retest reliability for the MINI-Kid is good for any disorder and .75–1.00 for individual disorders over 1–5 days (Sheehan et al., 2010)

Concerns

- Further validation is needed of the MINI-Screen with offender populations
- The MINI does not consider symptom severity, and thus may generate unnecessary referrals for treatment. The MINI does not assess cognitive impairment
- The MINI-Plus requires an average of 41 minutes to administer to offenders, which may inhibit broad use of the instrument with this population (Black et al., 2004)
- Although malingering, denial of symptoms, and other response sets are common problems in justice settings, the MINI is not able to detect the presence of these response sets
- The psychosis and major depression modules of the MINI-Plus can be

somewhat difficult and confusing to administer (Black et al., 2004)

- The MINI-Plus clinician-administered interview exhibits lower sensitivity for substance use disorder and dysthmia (42–52 percent), as determined by the SCID-I patient version. Further, MINI patient rated self-report diagnoses for many anxiety disorders, bulimia, and current/lifetime mania have low sensitivity (17–55 percent). Low sensitivity for the MINI clinician-administered interview was found for agoraphobia, simple phobia, and lifetime bulimia (46–63 percent), as determined by the CIDI
- Agreement between the clinician administered MINI and the SCID-I was low for many current/lifetime anxiety disorders, current psychotic disorders, current/lifetime substance use disorder, and dysthymia (kappas range 43–67 percent)
- Agreement between the clinician administered MINI and CIDI was low for many anxiety disorders, bulimia, and current/lifetime manic diagnoses (kappas range 43–68 percent; Sheehan et al., 1998)
- The MINI-Kid has poor sensitivity for current/lifetime psychotic disorder, major depressive disorder, dysthymia and panic disorders (43–64 percent; Sheehan et al., 2010), as determined by the K-SADS-PL

Availability and Cost

The MINI is available in paper and computerized versions. The paper form may be downloaded free of charge and used, once permission is provided by the author. A computerized version may be ordered for $295 or more, depending upon the version. The following website can be accessed to contact the author for permission to use the MINI or to obtain more information about the MINI 6.0, eMINI 6.0 (computerized version) and Dolphin EDC (MINI administered via internet browser): http://medical-outcomes.com/index/mini

The MINI Plus 5.0 can be downloaded for free at the following location: http://www.mdpu.ca/documents/mini.pdf

Psychiatric Research Interview for Substance and Mental Disorders (PRISM)

The PRISM is a semi-structured interview designed to diagnose psychopathology among substance-involved people. The instrument requires approximately 90 minutes to administer. As a result of the increasing recognition of the relevance of CODs, DSM-IV and DSM-5 emphasize the importance of distinguishing between substance-induced psychiatric symptoms related to active use and withdrawal and "primary" mental disorders (Samet, Nunes, & Hasin, 2004). Since specific guidelines for these diagnostic decisions did not exist prior to DSM-IV, in the past there have been problems with the reliability and validity of mental health diagnoses among people with substance use disorders. The PRISM examines current and lifetime substance use, mental disorders, and borderline and antisocial personality disorders. The substance use sections are presented prior to other diagnostic sections. Therefore, the interviewer has the substance use history information available when assessing mental disorders.

A computerized version of the PRISM (PRISM-CV-IV) is also available. The PRISM-CV-IV reviews the consistency of respondents' answers, and incorporates skip logic, reducing administration time to approximately 70 minutes (Hasin, Samet, Nunes, Mateseoane, & Waxman, 2006). A diagnostic report is produced to assist with scoring and interpretation. Differences between the paper and computerized version of the PRISM include use of a question format (e.g., multiple questions in the paper version are presented as individual questions in the computerized version). The order of modules is also different in the paper and computerized versions. Additional modules in the computerized version include nicotine use, suicidality

assessment, ADHD, and Pathological Gambling. The PRISM paper version is no longer supported by the PRISM website.

Positive Features

- The instrument distinguishes between primary and substance-induced disorders
- The PRISM was developed using a racially/ethnically diverse sample
- A Spanish version of the PRISM is available and appears to have some advantages over the Spanish version of the SCID in diagnosing major depression and borderline personality disorders among substance-involved people (Torrens, Serrano, Astals, Pérez-Domínguez, & Martín-Santos, 2004)
- The PRISM addresses the problem of diagnosing depression among people with substance use disorders
- The PRISM-CV has been widely used in both mental health and general medical settings
- Severity measures, consisting of a continuous rating of the number of symptoms present, are provided for some mental disorders, such as major depressive disorder and substance use disorders
- The PRISM has been used with several populations that have CODs (Coombes & Wratten, 2007; Hasin et al., 2002; Vergara-Moragues et al., 2012), with individuals who are homeless (Caton et al., 2005), and with offenders (Kravitz, Cavanaugh, & Rigsbee, 2002)
- Among substance-involved populations, the PRISM exhibits good agreement with DSM-IV diagnoses for current and lifetime diagnoses (kappas range .62–.82; Hasin et al., 2006)
- Among people with substance use disorders, the PRISM demonstrates good reliability for agreement in severity across most types of disorders, including both current and lifetime disorders (Hasin et al., 2006)
- Among people with substance use disorders, the PRISM shows adequate agreement with DSM-IV diagnoses of current and lifetime major depressive disorder and manic episodes, psychotic disorders, eating disorders, and personality disorders (Hasin et al., 2006)
- The PRISM has excellent reliability in diagnosing major depression (Hasin, Samet, Nunes, Mateseoane, & Waxman, 2006)

Concerns

- The PRISM interview must be administered by a trained clinician
- The PRISM website no longer supports the paper instrument services, such as data entry or diagnostic programs for scoring and interpretation
- The PRISM has not been widely used or tested in criminal justice populations
- Agreement with DSM-IV diagnoses of many substance use disorders has been found to be low in some samples (Hasin et al., 2006)
- Reliability for the PRISM severity of stimulant disorder is low, as determined by symptoms counts on the DSM-IV for both current and lifetime disorder (ICCs range .55–.64; Hasin et al., 2006)
- The PRISM's anxiety disorders module does not have good reliability for primary or substance-induced anxiety disorders (kappa = .57), nor dysthymic disorder (kappa = .36; Hasin et al., 2006)

Availability and Cost

The author of the PRISM maintains a website (http://www.columbia.edu/~dsh2/prism/) containing information regarding computer software related to the instrument. The site also contains information regarding the PRISM's psychometric properties and available training.

The training manual for the PRISM is available at the following location: http://www.columbia.edu/~dsh2/prism/files/PRISMman266.pdf

The PRISM-CV-IV is available for purchase and includes all software required for administration, scoring, and interpretation. PRISM administration does not require the software, but it is recommended that a license be purchased from Blaise ® Licensing. Information including cost (approximately $200) can be obtained by requesting a software quote through the following site: https://www.westat.com/our-work/information-systems/blaise percentC2 percentAE-distribution-training/blaise-licensing-ordering

The PRISM-CV-IV software package includes the interview protocol, a codebook that defines interview questions and diagnostic variables, a manual that provides diagnostic information for scoring and interpretation of interviews, a user guide, and information on how to export data to other statistical software programs. The cost of this package is $1,800.

Training and certification for administration of the PRISM-CV-IV is available. The cost of training workshops is $3,000 and certification costs are $200.

Paper instruments including the training manual for scoring and interpretation are available upon request by sending email correspondence to the following address: AivadyaC@nyspi.columbia.edu

Psychiatric Diagnostic Screening Questionnaire (PDSQ)

The PDSQ (Zimmerman & Mattia 2001b) is a 126-item self-administered instrument that assesses 13 of the most common DSM-IV mental disorders in outpatient mental health settings. The instrument was designed to assess current and recent symptomatology and to provide background information prior to providing a more extensive diagnostic evaluation. The PDSQ examines five areas, including eating disorders, mood disorders, anxiety disorders, substance use disorders, and somatoform disorders. The PDSQ also includes a six-item screen for psychosis. The instrument has undergone several iterations to enhance the reliability and validity, and indices of mania, dysthymic disorder, and anorexia were eliminated from the instrument due to poor psychometric features. At recommended cut-off scores, the PDSQ has sensitivity of greater than 90 percent for major depressive disorder, obsessive-compulsive disorder, PTSD, generalized anxiety disorder (GAD), panic/agoraphobia/social phobia, alcohol use disorders, and bulimia or somatoform disorders (Zimmerman, 2002; Zimmerman & Mattia, 2001a).

Positive Features

- The PDSQ requires only 15 minutes to administer, yet reviews a range of mental disorders
- The PDSQ was developed to be aligned with DSM diagnostic classifications
- The PDSQ has been used extensively with populations that have CODs and may assist in detecting disorders that are missed during unstructured clinical evaluations
- Cut-off scores were chosen to optimize sensitivity (> 90 percent; Zimmerman & Mattia, 2001a)
- The PDSQ has been used to diagnose mental disorders in justice settings (Stuart, Moore, Gordon, Ramsey, & Kahler, 2006; Swogger, Walsh, Houston, Cashman-Brown, & Connor, 2010; Weitzel, Nochajski, Coffey, & Farrell, 2007) and among people with substance use disorders (Simmons, Lehmann, & Cobb, 2008; Weitzel et al., 2007)
- PDSQ subscales related to depression are correlated with victimization of women and PTSD among women who are arrested for domestic violence (Stuart et al., 2006)
- Among offenders, the PDSQ subscales of GAD and PTSD are correlated with impulsive aggression (Swogger et al., 2010)
- The PDSQ results in a 42 percent rate of referral for further mental health evaluation among drug offenders, a rate similar to those referred for evaluation in other

substance-involved populations (Harris & Edlund, 2005; Watkins et al., 2004; Weitzel et al., 2007)

- The PDSQ has a low false positive rate in identifying Axis I disorders (30 percent; Zimmerman & Chelminski, 2006). Among psychiatric outpatients, the AUC for the PDSQ is good for those with and without diagnosed substance use disorders (.83 and .86 respectively) as determined by the SCID-I, across a range of disorders (Zimmerman, 2008; Zimmerman, Sheeran, Chelminski, & Young, 2004)
- Among psychiatric outpatients with substance use disorders, the PDSQ has good sensitivity (92 percent) and adequate specificity (63 percent) in identifying co-occurring mental disorders (Zimmerman, 2008; Zimmerman & Chelminski, 2006; Zimmerman et al., 2004)
- The PDSQ has good to excellent internal consistency (alphas ≥ .80 for 12 out of 13 subscales); test-retest reliability over two weeks (r score ≥ .80 for nine subscales, mean r score = .83); and discriminant, convergent, and concurrent validity (Zimmerman & Mattia, 2001a)

Concerns

- The validity of the PDSQ has not been widely studied in justice-involved populations for the diagnosis of mental disorders
- Various cut-off scores are recommended to achieve optimal sensitivity for mental disorders, which may lead to difficulties in scoring and interpreting results
- The PDSQ's alcohol and drug subscales do not distinguish between levels of substance use severity (Stuart et al., 2006)
- The PDSQ has low specificity for generalized anxiety disorder, obsessive-compulsive disorder, social phobia, and PTSD among people who are diagnosed with substance use disorders, as determined by the SCID-I (Zimmerman, 2008; Zimmerman et al., 2004)
- Positive predictive values for the PDSQ vary widely across mental disorders, indicating that some individuals may not be correctly diagnosed as having a disorder (Zimmerman, 2008; Zimmerman & Chelminski, 2006)
- The sensitivity of the PDSQ's psychosis subscale is not particularly high (Zimmerman, 2008; Zimmerman & Chelminksi, 2006; Zimmerman & Mattia, 2001a)
- No current PDSQ validity indices are available for mania, dysthymic disorder, or anorexia

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 $130 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).

Schedule of Affective Disorders and Schizophrenia–Third Edition (SADS)

The SADS is a semi-structured interview designed for use by trained clinicians to evaluate current and lifetime affective and psychotic disorders (Endicott & Spitzer, 1978). The instrument predates the SCID and offers specified probes for diagnostic criteria. The SADS includes Part I (Current) and Part II (Lifetime). Part I assesses current episodes, particularly the most severe period of the current episode. The SADS also examines six graduated levels of symptoms experienced, ranging from "not at all" to "extreme." Part II of the SADS reviews lifetime history of symptoms and episodes of the disorders and features two graduated levels of symptoms experienced ("presence" or "absence"). Several alternate versions of the SADS have also been developed. For example, the SADS-L is similar to Part II of the SADS in that it provides

a description of lifetime symptoms and dedicates very little time to current symptoms. The 45-item SADS-C examines current symptoms and changes in these symptoms. The global assessment scale of the SADS-I describes symptoms experienced over particular intervals of time following the initial SADS-L interview.

Positive Features

- The SADS has been found to be more effective than the DIS in diagnosing depressive disorders (Hasin & Grant, 1987)
- Interrater reliability is excellent for current disorders and is good for past disorders
- The SADS has been translated into several languages
- The instrument examines symptom severity and ancillary symptoms that are related to, but not part of, formal diagnostic criteria
- The SADS has been used in justice settings to diagnose mental disorders (Blackburn & Coid, 1998; Hodgins, Lapalme, & Toupin, 1999) and has been found to be effective in these settings (Rogers, Sewell, Ustad, Reinhardt, & Edwards, 1995; Rogers, Jackson, Salekin, & Neumann, 2003)
- The SADS is useful in inpatient, outpatient, and primary health care settings for diagnosing CODs and providing referral to services (Rogers, Jackson & Cashel, 2004)
- The SADS has adequate concurrent validity for mental disorders when compared with other diagnostic interview instruments (Farmer et al., 1993; Rogers et al., 2004; Hesselbrock, Stabenau, Hesselbrock, Mirkin, & Meyer, 1982)
- The SADS-C has good reliability in diagnosing mental disorders (McDonald-Scott & Endicott, 1984)
- The SADS-C subscales of schizophrenia, depression, and bipolar disorder are significantly correlated with similar scales on the Referral Decision Scale (Rogers, Sewell et al., 1995), and other studies provide evidence of concurrent validity of the SADS-C (Johnson, Magaro, & Stern, 1986)
- Within justice settings, the SADS-C shows good interrater reliability for symptoms and subscales (ICC = .92, range .94–.97; Rogers et al., 2003) in both treatment seeking and emergency care settings
- Across multiple studies, the SADS exhibits good interrater reliability for symptom ratings and diagnosis (Andreasen et al., 1982; Endicott, & Spitzer, 1978; Keller et al., 1981; Rogers, Sewell et al., 1995)
- The SADS's test-retest reliability is moderate to high (McDonald-Scott & Endicott, 1984; Rapp, Parisi, Walsh, & Wallace,1988) when the elapsed time between administrations is less than 6 months

Concerns

- The SADS was developed concurrently with the DSM-III and does not use DSM-IV or DSM-5 terminology or classification systems
- There is poor agreement between the SADS and the DIS in diagnosing depression among individuals with substance use problems (Hasin & Grant, 1987)
- The SADS does not adequately address all substance use disorders, and thus, other interviews such as SCID may be preferred (Rogers, 2001)
- The SADS has not been used extensively in justice settings
- The SADS is rather lengthy and complex to administer and requires clinical judgment
- Significant training is required for administration and scoring of the SADS
- The instrument is not very sensitive to response styles, and participants can fake positive symptoms of disorders. Research has examined the potential use of some SADS-C subscales to detect malingering (Rogers et al., 2003)

- The SADS provides limited breadth of coverage, with a focus on evidence of affective and psychotic disorders
- The SADS is not recommended for assessment of personality disorders (Rogers, 2001)

Availability and Cost

A description of the SADS can be found in the following article: Endicott, J., & Spitzer, R. L. (1978). A diagnostic interview: The Schedule of Affective Disorders and Schizophrenia. *Archives of General Psychiatry, 35*, 837–844.

This instrument is no longer in print and thus copies of the instrument may be difficult to obtain.

Structured Clinical Interview for DSM-IV (SCID-IV)

The SCID is a semi-structured psychological assessment interview developed for administration by trained clinicians (First, Spitzer, Gibbon, & Williams, 1996). The SCID-I is one of the most widely used structured interview instruments developed to diagnose DSM disorders and is considered to be the "gold standard" for diagnostic assessment (Shear et al., 2000). The SCID-I obtains diagnoses for all mental disorders, using the DSM criteria. Standard threshold questions are provided and the administrator may reword questions to clarify them, as needed. The Substance Use Disorders module identifies lifetime and past 30-day diagnoses for alcohol and other drugs. The SCID-IV also differentiates between different levels of severity of substance use disorders. A separate instrument (SCID-II) examines Axis II Personality Disorders and is published separately.

Both research (SCID-RV) and clinical versions (SCID-CV) of the SCID-I and II are available. The clinical version is shorter (45–90 minutes) and examines disorders frequently seen in clinical settings (First, Spitzer, Gibbon, & Williams, 2001), while excluding most of the subtypes, severity, and course specifiers included in the research version. Some disorders are not fully evaluated but instead are assessed briefly at the end of the SCID administration (e.g., social and specific phobia, generalized anxiety disorder, eating disorders, hypochondriasis). The full SCID-I Research Version examines the mental disorders. The Research Version requires approximately 1.5–2 hours to administer and 10 minutes to score.

The SCID-RV and SCID-CV for DSM-5 are now available, in addition to user guides for these instruments. These instruments are available from the American Psychiatric Publishing Inc. (see "Availability and Cost"). Revisions are also underway for the SCID-II, which will be renamed the "SCID for Personality Disorders" (SCID-PD).

Positive Features

- Diagnoses are made according to DSM-IV, DSM-IV TR, or DSM-5 criteria
- The SCID has been translated into several languages. Several foreign language versions have been shown to have good psychometric properties (Lobbestael, Leurgans & Arntz, 2011; Schneider et al., 2004)
- Computer-assisted interview versions of the SCID (SCID-CV) are available, including the research version. A shorter, computer-administered self-report screening version of the SCID is also available. However, this latter version does not yield definitive diagnoses but rather diagnostic impressions that should be confirmed through use of a SCID interview or full clinical evaluation
- The instrument has been used with psychiatric, medical, nonsymptomatic adults in the community and justice populations (Cohen et al., 2002; Dolan & Blackburn, 2006; Morgan, Fisher, Duan, Mandracchia, & Murray, 2010; First et al., 2001; Peters. Greenbaum, Edens, Carter, and Ortiz, 1998; Peters et al., 2000)
- SCID diagnoses have been found to be more accurate and more comprehensive than unstructured clinical interviews (Basco et al., 2000; Kranzler et al., 1995)

- The SCID has been used to assess CODs, including treatment-seeking individuals who have substance use disorders (Kidorf et al., 2004)
- In a community sample, the SCID for Axis II disorders shows adequate interrater reliability for diagnoses (kappas range .85–.95) in addition to adequate agreement for the presence of individual traits related to mental disorders (ICCs range .87–.99). The self-report SCID-II demonstrates good interrater reliability for the diagnosis of the personality disorders (kappas range .66–.99; Farmer & Chapman, 2002)
- Peters et al. (1998) examined the use of the SCID among correctional populations using DSM-IV guidelines. Kappas were moderately high for alcohol disorders (current diagnosis, .80; lifetime diagnosis, .78) and varied considerably for drug use disorders (current diagnosis, .48–1.00; lifetime diagnosis, .04–1.00), although these were generally quite high
- The SCID shows good interrater reliability in people receiving outpatient treatment across mental disorders (Zimmerman & Mattia, 1999a) and for both lifetime and past month alcohol and drug disorders among offenders (Peters et al., 2000)
- The internal consistency of the SCID-II is good, with alphas ranging .71–.94 (Maffei et al., 1997)

Concerns

- The SCID was designed for use by a trained clinician at the masters or doctoral level, although in research settings, it has also been used by bachelors-level technicians with extensive training. Significant training is required for both administration and scoring of the SCID
- Administration of the SCID I and II may each require more than 2 hours for individuals who have multiple diagnoses. The Psychoactive Substance Use Disorders module requires 30–60 minutes, when administered separately
- For people with cognitive impairment or psychotic symptoms, the SCID may need to be administered across several sessions
- Clinical judgment is required to determine whether symptoms are present for a particular disorder
- An eighth-grade reading level is required for the SCID
- The SCID provides a dichotomous decision (yes/no) regarding diagnoses, and it does not provide subthreshold diagnoses or take into account symptoms that may be experienced along a continuum
- The SCID is quite costly to purchase

Availability and Cost

The SCID is available for purchase from American Psychiatric Publishing, Inc., 1400 Street, N.W., Washington, DC 20005, at the following site: http://www.appi.org/home/search-results?FindMeThis=SCID

Available materials include SCID user's guides, administration booklets, and score sheets. The Research Version of the SCID can be obtained by contacting Biometrics Research at (212) 960-5524.

The user's guide and administration booklet cost approximately $80 for either the SCID-I or SCID-II. A packet of SCID score sheets costs approximately $80.

The SCID-5 products can be purchased at the following site: https://www.appi.org/products/structured-clinical-interview-for-dsm-5-scid-5

Recommendations for Assessment and Diagnosis of CODs

Information describing assessment and diagnostic instruments related to co-occurring mental and substance use disorders is based on a critical review of the instruments and research examining their efficacy. Key considerations in recommending instruments are based upon empirical evidence supporting both the reliability and validity of the instrument, relative cost of the instrument, ease of administration, and use

within justice settings. Although summaries of instruments are based on DSM-IV criteria, instruments recommendations are those that align more closely with DSM-5, allowing for a more seamless transition from DSM-IV to DSM-5. Recommendations for assessment and diagnosis of co-occurring mental and substance use disorders include instruments that provide comprehensive examination of multiple disorders and related biopsychosocial problems. The following instruments are recommended:

1. The Alcohol Use Disorders and Associated Disabilities Interview (AUDADIS-IV), which provides a comprehensive assessment and examines a range of co-occurring substance use and mental health problems, including personality disorders and psychosocial risk factors.

(or)

2. The Mini International Neuropsychiatric Interview (MINI) or the Structured Clinical Interview (SCID), which address a full range of co-occurring mental health and substance use disorders and provide a diagnostic impression of multiple disorders.

Each instrument requires between 45-120 minutes to administer, dependent on the symptom presentation and particular problems that are selected for assessment. The measures can be administered in their entirety, or specific modules can be administered that are tailored to the individual's assessment needs and set of symptoms. The different options provided here for assessment and diagnosis of co-occurring disorders may be appealing dependent on the specific needs in a particular justice setting. The MINI and SCID provide diagnosis of the full set of disorders, while the AUDADIS provides a comprehensive assessment of the disorders and a review of related biopsychosocial problems. These instruments should be administered by trained clinicians who are licensed, certified, or otherwise credentialed in assessing and diagnosing CODs and related psychosocial problems.

Substance Abuse and Mental Health Services Administration. (2019, June).
Screening and Assessment of Co-occurring Disorders in the Justice System.
HHS Publication No. PEP19-SCREEN-CODJS.
Rockville, MD: Substance Abuse and Mental Health Services Administration.