



Development of a computerized adaptive substance use disorder scale for screening, measurement and diagnosis – The CAT-SUD-E



Leslie A. Hulvershorn^{a,*}, Zachary W. Adams^a, Michael P. Smoker^a, Matthew C. Aalsma^c, Robert D. Gibbons^b

^a Department of Psychiatry, Pediatric Care Center, Indiana University, 1002 Wishard Blvd., Suite 4110, Indianapolis, IN 46202, United States

^b Departments of Medicine and Public Health Sciences, The University of Chicago Biological Sciences, Chicago, IL, United States

^c Department of Pediatrics, Section of Adolescent Medicine, Indiana University School of Medicine, 410 West 10th Street Suite 2000, Indianapolis, IN, 46202, United States

ARTICLE INFO

Keywords:

Substance use disorder
Item response theory
Computerized adaptive testing
Mental health
Validation
DSM-5

ABSTRACT

Introduction: The Computerized Adaptive Test for Substance Use Disorder (CAT-SUD), an adaptive test based on multidimensional item response theory, has been expanded to include 7 specific Diagnostic and Statistical Manual, 5th edition (DSM-5) defined SUDs. Initial testing of the new measure, the CAT-SUD expanded (CAT-SUD-E) is reported here.

Methods: 275 Community-dwelling adults (ages 18–68) responded to public and social-media advertisements. Participants virtually completed both the CAT-SUD-E and the Structured Clinical Interview for DSM-5, Research Version (SCID) to assess the validity of the CAT-SUD-E in determining whether participants met criteria for specific DSM-5 SUDs. Diagnostic classifications were based on 7 SUDs, each with 5 items, for current and lifetime SUDs.

Results: For SCID-based presence of any lifetime SUD, predictions based on the overall CAT-SUD-E diagnosis and severity score were AUC=0.92, 95% CI = 0.88, 0.95 for current and AUC=0.94, 95% CI = 0.91, 0.97 for lifetime. For individual diagnoses, classification accuracy for current SUDs ranged from an AUC=0.76 for alcohol to AUC=0.92 for nicotine/tobacco. Classification accuracy for lifetime SUDs ranged from an AUC=0.81 for hallucinogens to AUC=0.96 for stimulants. Median CAT-SUD-E completion time was under 4 min.

Conclusions: The CAT-SUD-E quickly produces similar results as lengthy structured clinical interviews for overall SUD and substance-specific SUDs, with high precision and accuracy, through a combination of fixed-item responses for diagnostic classification and adaptive SUD severity measurement. The CAT-SUD-E harmonizes information from mental health, trauma, social support and traditional SUD items to provide a more complete characterization of SUD and provides both diagnostic classification and severity measurement.

1. Introduction

Substance use disorders (SUDs) are a public health emergency (Haffajee and Frank, 2018), where the majority of people with SUDs do not receive treatment for their SUDs and related mental health disorders (Creedon and Lê Cook, 2016). Identifying those in need is a major barrier to receiving treatment (Priester et al., 2016). SUD identification and treatment is predicated on accurate diagnoses, monitoring changes in risk and severity over time, and effective, timely interventions, targeted at the identified SUDs (Substance Abuse and Mental Health Services Administration (US), 2016). In addition to initial detection of risk, instruments are needed that also deliver efficient, accurate quantification of non-negligible risk to assist in decision making and allocation of

resources across diverse health care settings (e.g. emergency, outpatient, inpatient, behavioral health) (Gibbons et al., 2020).

In response to the need for effective detection and measurement of SUDs, and enabling measurement-based care, Gibbons et al. (2020) developed the first computerized adaptive test (CAT) based on multidimensional item response theory (MIRT) for SUD: the CAT-SUD. The CAT-SUD provides a psychometric harmonization between SUD, depression, anxiety, trauma, social isolation, functional impairment, and risk-taking behavior symptom domains, providing a balanced multidimensional view of SUD. An item-bank of 252 items drawn from five subdomains (1) SUD, (2) psychological disorders, (3) risky behavior, (4) functional impairment, and (5) social support was developed. Calibration of the item bank using a bifactor IRT model (Gibbons and Hedeker, 1992;

* Corresponding author.

E-mail address: lhulvers@iupui.edu (L.A. Hulvershorn).

<https://doi.org/10.1016/j.dadr.2022.100047>

Received 13 September 2021; Received in revised form 4 March 2022; Accepted 18 March 2022

2772-7246/© 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Gibbons et al., 2007) revealed excellent fit to the item bank, with 168 items having primary factor loadings >0.4 , which were retained in the CAT-SUD. Adaptive administration required an average of 11 items (range: 4–26), in less than 2 min, producing a 94% reduction in participant burden, yet maintained a correlation of $r = 0.91$ with the total 168-item bank score. The CAT-SUD was validated against structured clinical interviews based on the Composite International Diagnostic Interview (CIDI) SUD diagnosis (AUC=0.85, 95% confidence interval (CI) = 0.75, 0.95). The scale score is presented on a 0–100 point scale with 5 points of precision. Severity thresholds of low- (<50), intermediate- (50–70), and high-risk (>70) were derived from discrete points on the receiver operator characteristic (ROC) curve that maximize sensitivity between the low-risk and intermediate-risk categories and maximize specificity between the intermediate-risk and high-risk categories. The thresholds resulted in rates of 4, 22 and 50% for CIDI SUD diagnoses and 11, 47 and 90% for self-reported alcohol or drug use for low-, intermediate- and high-risk CAT-SUD categories.

The CAT-SUD represents a psychometric advance in the measurement of SUD severity and diagnostic screening of an overall SUD diagnosis. The original CAT-SUD also collects self-reports of 5 categories (opiates/analgesics, alcohol, cocaine/amphetamines, heroin/methadone, sedatives/hypnotics/tranquilizers/barbiturates) of past-month substance use frequency. However, it provides neither substance-specific diagnoses nor specification of disorder timeframe (current versus past/lifetime). Such information is critical to treatment planning, especially in busy clinical settings where comprehensive diagnostic evaluations are often limited by time and resource constraints. As noted in the original publication (Gibbons et al., 2020), an expansion of the CAT-SUD – the CAT-SUD-expanded (CAT-SUD-E) – was developed to bridge this gap and provide individual substance-specific SUD diagnoses. With the CAT-SUD-E, we consider the following 7 DSM-5 (American Psychiatric Association, 2013) SUDs at the present time and over the participants' lifetimes: (1) alcohol, (2) cannabis, (3) opioids, (4) stimulants, (5) sedatives, (6) hallucinogens, (7) nicotine/tobacco. The CAT-SUD-E extends the CAT-SUD by adding a separate self-reported branching logic interview based on DSM-5 criteria for specific drugs of abuse, current and in the past. The original CAT-SUD severity score and the individual drug-specific DSM-5 criteria are then used in combination to provide SUD diagnostic classifications. We call this the CAT-SUD-E, which provides both SUD diagnostic classifications and severity scores based on the original CAT-SUD. The purpose of this paper is to report on the expansion of the CAT-SUD to the CAT-SUD-E and the direct comparison, in adults, of the CAT-SUD-E against a gold standard diagnostic assessment, the Structured Clinical Interview for DSM-5, Research Version (SCID) (First et al., 2015).

2. Materials and methods

2.1. Design

This study was conducted in compliance with the ethical principles of the Declaration of Helsinki and the International Conference on Harmonization's Good Clinical Practices Guidelines from April 2020 to April 2021. The Institutional Review Board at Indiana University approved the study, and individuals provided verbal informed consent prior to initiation of study procedures. The design of this prospective study involved administration of both the self-reported CAT-SUD-E and a clinician-conducted interview using the SUD section of the SCID. The CAT-SUD-E assessed both present (past 30 days) and lifetime (prior to the past 30 days) SUDs. The SCID assessed present (past 12 months) and past (prior to the past 12 months) SUDs. As the SCID does not typically assess for nicotine/tobacco use disorder criteria, a new set of items was created by adapting DSM-5 SUD criteria for nicotine and tobacco into a format similar to SCID SUD items for other substances, for the SCID interviews. Nineteen participants completed interviews prior to the realization of a need for, and addition of, nicotine-related SCID items;

therefore, all CAT-SUD-E responses, except for nicotine, were included for these individuals. The order of administration for the CAT-SUD-E and SCID was randomized across participants, and the tests were administered on the same day in 87% of cases.

The sample was recruited from the community using multiple strategies, including physical and online advertisement, referrals from community mental health centers, and word-of-mouth referrals from previous participants. Eligibility criteria included being 18 years or older, being fluent in English, and having access to an electronic device permitting the use of Zoom and CAT-SUD-E applications, and virtually all individuals satisfied these criteria. Attempts were made to recruit participants with a diverse set of substance use histories. Mental health centers were a key recruitment and referral source, which likely contributed to enrollment of a relatively high proportion of participants who met criteria for one or more SUDs. Initial phone screens were used, in part, to help ensure a sample likely enriched in SUD diagnoses, with approximately 70% of participants having endorsed at least one negative consequence associated with substance use, while the remaining 30% of participants having denied any such consequences during their screen. Once the target number of participants yielding a SCID-based SUD-positive diagnosis was reached, recruitment was narrowed to individuals with little to no substance use experience. All interviews were conducted virtually using the secure Zoom Health tele-videoconferencing platform. SCID assessors were five experienced Masters- or Doctoral-level behavioral health clinicians working under the supervision of a licensed psychologist or psychiatrist. They were blind to the results of the CAT-SUD-E. Participants were compensated with a \$100 gift card and had the option of sharing the results of their SCID with their treating clinicians, if relevant, or viewing it themselves. Of 287 participants, 12 were not included in the analytical sample ($n = 275$) for at least one of the following reasons: incomplete SCID, incorrect CAT-SUD version (not Expanded), incorrect CAT-SUD-E timeframe.

2.2. The CAT-SUD

The calibration and validation of the CAT-SUD was reported previously (Gibbons et al., 2020). Briefly, the CAT-SUD was developed based on the following steps. First, an item bank was constructed to cover the primary domain and each of the subdomains of importance. Ordinal (Likert-type) response categories are more informative in IRT applications than binary response categories, so most items use a 5-point Likert-type scale. Second, the item bank (either in its complete form or reduced into subsets based on a balanced incomplete block design, see Gibbons et al., 2012) was administered to a large sample of participants. Third, the item-response data were then calibrated using an MIRT model to preserve their multidimensionality. The item bank was then restricted to those items that have strong loadings (>0.4) on the primary dimension, which provided a cross-walk between the primary domain and the subdomains, preserving the multidimensionality of the construct into a single-valued index of the construct. Fourth, using the item parameter estimates, adaptive testing was simulated from the complete item-response data, to optimize tuning parameters of the CAT (e.g. termination criteria, probability of taking the next optimal item or the second optimal item to spread use of items across the entire item bank). This step also allowed us to create a CAT that minimized participant burden while maximizing the correlation between the CAT and the full bank administration. Fifth, the live CAT was then developed into a web-app and administered to a new validation sample of participants who also received a structured clinical interview (CIDI). As the CAT produces a continuous severity score, a logistic regression was used to relate the CAT score(s) to the binary clinician-based diagnosis, generating an ROC curve which can be characterized by the area under the ROC curve (AUC) and its 95% confidence interval. Cross validation was conducted using split samples or n-fold cross-validation where the model is fit on n-1 subsets of the sample and applied to the omitted subsample not used in deriving the classification function.

The process is repeated until all of the sample are used in the validation. Hosmer and Lemeshow (2000) provided the following classification system for the AUC: $0.7 \leq \text{AUC} < 0.8$ = 'Acceptable discrimination'; $0.8 \leq \text{AUC} < 0.9$ = 'Excellent discrimination'; $\text{AUC} \geq 0.9$ = 'Outstanding discrimination'. As noted above, applying this methodology to the development of the CAT-SUD produced an adaptive test that extracted the information from a 168-item bank using an average of 11 items that can be administered in approximately 2 min. Validation revealed an AUC in the "excellent discrimination" range (Gibbons et al., 2020).

2.3. The items

The CAT-SUD-E extends the CAT-SUD by adding a series of branching logic questions about drug use and DSM-5 criteria for specific drugs of abuse, currently and in the past. The item bank for the adaptive portion of the CAT-SUD-E is identical to the original 168 items in the CAT-SUD (Gibbons et al., 2020). All items were administered electronically (i.e., via smartphone, computer or tablet) in writing or were read to participants digitally from the device, if they desired. The diagnostic interview portion began with determining if the participant was over 18 years of age (an adolescent version of the CAT-SUD-E is still being validated). Thereafter, questions regarding hospitalizations in drug treatment centers and probation in the past 3 months were asked, as hospitalization and probation could increase SUD abstinence and, thus, impact study findings. As above, the time frame of the questions (diagnostic and adaptive severity measurement) were asked both over an interval of the previous 30 days and across the participants' lifetimes.

Participants were then asked if they had used any of the following substances during the selected time-frame (and then again, separately for lifetime use): alcohol, cannabis/marijuana/spice, nicotine (vaping, smoking, chewing), opioids (examples: morphine, Percocet, dilaudid, Vicodin, Oxycotin, methadone, heroin), cocaine/crack, methamphetamines/amphetamines (examples: Adderall, meth), hallucinogens, and sedatives (examples: benzodiazepines like Xanax, Klonopin or Ambien). For each substance selected for each time-frame (e.g. past 30 days and lifetime), the following 5 questions, based on DSM-5 criteria, were asked (illustrated here for the past 30 days and alcohol), with no/yes response categories:

- 1 When I stop using alcohol I get withdrawal symptoms (for example, sweating, heart racing, shakiness, seizures, nausea/vomiting, hallucinations).
- 2 In the past month have you spent a lot of time using, obtaining, thinking about or craving alcohol?
- 3 In the past month have you or someone else wanted you to cut back on alcohol?
- 4 In the past month has alcohol caused problems for you or kept you from getting things done?
- 5 In the past month, have you used alcohol again and again, despite the harm it has caused?

These five items were selected based upon relative frequency of endorsement by people with SUDs in epidemiological studies and field trials (e.g., Hasin et al., 2013) and clinical experience of the authors.

2.4. Diagnostic scoring

For a current or lifetime diagnoses, yes to 2 or more of the above 5 diagnostic items is sufficient to classify the participant as meeting criteria for that particular SUD, consistent with minimum number of criteria for a mild SUD per DSM-5. Additionally, item 1 (withdrawal), by itself, is sufficient to generate a diagnosis for nicotine or alcohol. Affective withdrawal symptoms, such as anxiety and irritability, are present for most substances, but they are the primary symptoms for stimulants and cannabis and are often not recognized/appreciated by users, by the public, and to some extent by the medical field Katz et al., 2014; Livne et al., 2019; Pennay and Lee, 2011; Walker et al., 2019. In contrast, nicotine,

alcohol, benzodiazepines, and opioids all have more pronounced physiological withdrawal symptoms, which are more widely recognized, and in the case of alcohol (and rarely benzodiazepines) can be fatal Authier et al., 2009; Bayard et al., 2004; Brady et al., 2016; Kenny and Markou, 2001. Importantly, while withdrawal symptoms of benzodiazepines and opioids are readily recognizable, dependence on these substances (and subsequent withdrawal) can be produced by use as medically prescribed, unrelated to SUD Authier et al., 2009; Brady et al., 2016. Thus, only alcohol and nicotine withdrawal symptoms are both relatively easy to identify and not attributable to medical use, which is why endorsement of the withdrawal symptom was interpreted differently for alcohol and nicotine.

2.5. Statistical analysis

For the validation analysis, we used logistic regression to examine the association between the continuous CAT-SUD score (0–100-point scale, transformed from the underlying unit normal score for ease of interpretation by clinicians) and CAT-SUD-E diagnosis as predictors, and the SCID SUD diagnoses as outcomes. This was done separately for current and lifetime for the overall and individual SUD diagnoses. For each logistic regression model, we computed the probability of a SCID SUD diagnosis, and then computed the AUC of the ROC curve. AUCs were computed for current, and lifetime diagnoses overall and for each individual diagnosis. For each analysis, sensitivity was also reported at fixed specificity values of 0.8 and specificity of 0.9, and sensitivity, specificity and kappa were also computed at the point on the ROC curve of maximum classification accuracy.

Of note, the SCID refers to substance use within the past 12 months, whereas the CAT-SUD-E refers to the past 30 days, so our estimates of agreement represent a lower bound on what can be achieved using the same time-frames. In addition, The SCID diagnoses include the category stimulants, whereas the CAT-SUD-E includes the categories methamphetamine/amphetamine and cocaine. To this end, we created the stimulant category which included stimulants for the SCID as a single "stimulant" category and included methamphetamine/amphetamine and cocaine as "stimulants" for the CAT-SUD-E. This also places an upper bound on the computed level of agreement.

3. Results

3.1. Descriptive statistics

Table 1 displays sample demographic characteristics based on SCID diagnoses. While some participants declined to provide race/ethnicity values, the majority listed as "NA" in Table 1 were due to researchers inadvertently not collecting these measures at the outset of enrollment. Two hundred seventy-five participants (ages 18–68; 57% female) completed the CAT-SUD-E and a SCID DSM-5 diagnostic interview. Median CAT-SUD-E completion time was 3 min and 40 s (interquartile range from 1:59 to 5:53). Conversely, based on estimates from clinicians administering the SCID, the median SCID completion time was approximately 60 min, with completion times ranging from 15 to 105 min, depending on participant substance use history (none to extensive, respectively). No adverse events from undergoing the assessments were noted, other than one participant who felt that being asked about drugs of abuse triggered an urge to use, although they denied acting on those urges.

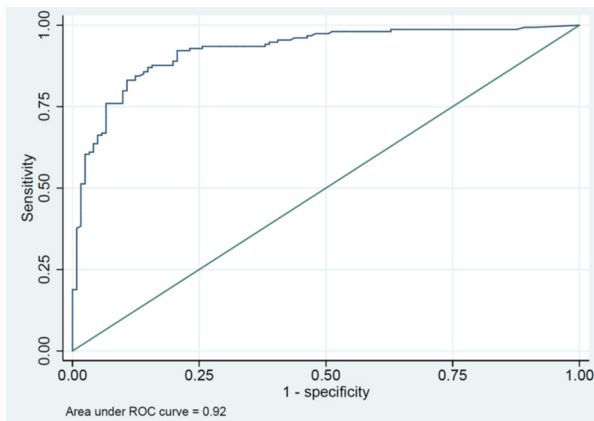
3.2. Overall diagnostic accuracy

For the overall current SUD SCID diagnosis, the CAT-SUD-E (continuous severity score and diagnostic screener) had $\text{AUC}=0.92$, 95% $\text{CI} = 0.88, 0.95$, which is in the "outstanding" range (see Fig. 1). Maximum classification accuracy was 87%, and at that point on the ROC

Table 1
Sample description.

Characteristic	Dx - (n = 79)		Dx + (n = 196)		All (n = 275)	
	M	(SD)	M	(SD)	M	(SD)
Age	32.2	(12.12)	36.1	(11.75)	34.9	(11.96)
Gender	<i>n</i>	(%)	<i>n</i>	(%)	<i>n</i>	(%)
Female	40	(50.6%)	116	(59.2%)	156	(56.7%)
Male	39	(49.4%)	79	(40.3%)	118	(42.9%)
Other	-	-	1	(0.5%)	1	(0.4%)
Race						
Native Am	1	(1.3%)	-	-	1	(0.4%)
Asian	16	(20.3%)	5	(2.6%)	21	(7.6%)
Black	5	(6.3%)	4	(2.0%)	9	(3.3%)
HI/Pac Island	-	-	-	-	-	-
White	46	(58.2%)	103	(52.6%)	149	(54.2%)
Multi	2	(2.5%)	10	(5.1%)	12	(4.4%)
N/A	9	(11.4%)	74	(37.8%)	83	(30.2%)
Ethnicity						
Hisp/Latinx	17	(21.5%)	5	(2.6%)	22	(8.0%)
Non	54	(68.4%)	114	(58.2%)	168	(61.1%)
N/A	8	(10.1%)	77	(39.3%)	85	(30.9%)

Dx - = Participants receiving no diagnoses from the SCID interview
 Dx + = Participants receiving ≥ 1 diagnosis from the SCID interview
 N/A = not asked or not provided

**Fig. 1.** ROC curve for current SUD.

curve, sensitivity was 0.92, specificity was 0.79, and kappa was 0.72. Increasing specificity to 0.90, yields sensitivity of 0.80.

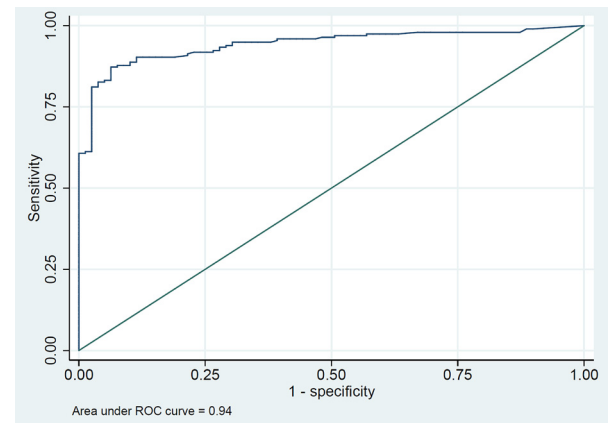
For overall lifetime SUD, AUC=0.94, 95% CI = 0.91, 0.97, is also in the “outstanding” range (see Fig. 2). Maximum classification accuracy was 90%, and at that point on the ROC curve, sensitivity was 0.90, specificity was 0.89, and kappa was 0.75. Increasing specificity to 0.90, yields sensitivity of 0.88.

Using just the CAT-SUD-E (diagnostic interview portion), apart from the CAT-SUD (severity score), yielded an AUC=0.82, 95% CI = 0.77, 0.88 (current SUD) and an AUC=0.88, 95% CI = 0.83, 0.92 (lifetime SUD).

To test for possible order effects (i.e., SCID first vs. CAT-SUD-E first), classification accuracy was computed. No differences were observed for either current (SCID first, AUC = 0.93, 95% CI = 0.89, 0.97; CAT-SUD-E first, AUC = 0.92, 95% CI = 0.87, 0.97) or lifetime (SCID first, AUC = 0.94, 95% CI = 0.90, 0.98; CAT-SUD-E first, AUC = 0.94, 95% CI = 0.91, 0.98) SUD.

3.3. Substance-specific diagnostic accuracy

Table 2 presents classification accuracy for the 7 specific substances for current SUD diagnoses. Classification accuracy ranged from acceptable for alcohol (AUC=0.77, 95% CI = 0.64, 0.80) and sedatives

**Fig. 2.** ROC curve for lifetime SUD.

(AUC=0.77, 95% CI = 0.55, 0.87), to excellent for cannabis (AUC=0.86, 95% CI = 0.74, 0.88), to outstanding for opioids (AUC=0.90, 95% CI = 0.79, 0.93), stimulants (AUC=0.91, 95% CI = 0.80, 0.92) and nicotine/tobacco (AUC=0.92, 95% CI = 0.91, 0.98). There were too few positive tests for hallucinogens to provide a reliable analysis. Table 2 reveals that maximum classification accuracy ranged from 84% for alcohol to 91% for stimulants. At the point of maximum classification accuracy, specificity ranged from 0.94 to 0.97, with sensitivity ranging from 0.35 for alcohol, to 0.80 for nicotine/tobacco. Kappa statistics for agreement ranged from 0.36 for alcohol to 0.76 for nicotine/tobacco.

Table 3 presents classification accuracy for the 7 specific substances for lifetime diagnoses. Classification accuracy ranged from excellent for hallucinogens (AUC=0.81, 95% CI = 0.69, 0.87), alcohol (AUC=0.85, 95% CI = 0.78, 0.89), cannabis (AUC=0.85, 95% CI = 0.79, 0.89), and sedatives (AUC=0.88, 95% CI = 0.79, 0.90), to outstanding for nicotine/tobacco (AUC=0.90, 95% CI = 0.844, 0.93), opioids (AUC=0.95, 95% CI = 0.91, 0.97), and stimulants (AUC=0.96, 95% CI = 0.91, 0.97). Table 3 reveals that maximum classification accuracy ranged from 79% for alcohol and cannabis to 93% for opioids. At the point of maximum classification accuracy, specificity ranged from 0.74 to 0.97, with sensitivity at 0.82 or greater with the exception of nicotine/tobacco 0.46 and sedatives 0.63. Kappa statistics for agreement ranged from 0.46 for hallucinogens to 0.84 for opioids and stimulants.

4. Discussion

Building on the development of the CAT-SUD, the CAT-SUD-E improves detection of overall SUD and substance-specific SUD diagnoses using a web-based modality that is efficient, scalable, and flexibly implemented. The CAT-SUD (severity score) was shown to have AUC=0.85 for a current SUD diagnosis, whereas the CAT-SUD-E improved classification accuracy to AUC=0.92 for current and AUC=0.94 for lifetime. Furthermore, AUC values dropped to 0.82 and 0.88, for current and lifetime, respectively, when using the CAT-SUD-E diagnostic interview alone. Results of the combination of the CAT-SUD-E severity score and specific diagnoses yielded outstanding classification accuracy for both current and lifetime assessments. Importantly, these metrics were achieved in under 4 min without the need for a trained clinician. This is especially important given the well-documented workforce shortages across behavioral health that may preclude providers' ability to implement comprehensive, diagnostic interviews in routine practice (University of Michigan Behavioral Health Workforce Research Center, 2018; Health Resources and Services Administration (HRSA), 2018). Kappa statistics for agreement between the assessments were 0.72 for current and 0.75 for lifetime disorders at the point of maximum classification accuracy. As a point of reference, reliability for alcohol use disorder for the DSM-5, the only SUD reported, was only 0.40 (Regier et al.,

Table 2
Substance-specific diagnostic accuracy - current.

Substance	AUC	Se at Sp=0.80	Se at Sp=0.90	MCA	Se at MCA	Sp at MCA	Pr(Dx) at MCA	Kappa at MCA
Alcohol	0.77	0.47	0.40	84%	0.35	0.96	0.51	0.36
Cannabis	0.86	0.68	0.58	85%	0.46	0.96	0.52	0.48
Opioid	0.90	0.76	0.62	89%	0.41	0.97	0.48	0.38
Stimulant	0.91	0.87	0.68	91%	0.62	0.97	0.52	0.67
Sedative	0.77	0.50	0.38	Too few	Too few	Too few	Too few	Too few
Hallucinogen	Too few	Too few	Too few	Too few	Too few	Too few	Too few	Too few
Nicotine/Tobacco	0.92	0.86	0.83	89%	0.80	0.94	0.77	0.76

Model includes current CAT Dx and current severity score

MCA=Maximum classification accuracy, Se=sensitivity, Sp=specificity

For CAT-SUD-E stimulant = cocaine and amphetamine

Table 3
Substance-specific diagnostic accuracy – lifetime.

Substance	AUC	Se at Sp=0.80	Se at Sp=0.90	MCA	Se at MCA	Sp at MCA	Pr(Dx) at MCA	Kappa at MCA
Alcohol	0.85	0.77	0.62	79%	0.82	0.74	0.41	0.55
Cannabis	0.85	0.76	0.63	79%	0.82	0.76	0.30	0.57
Opioid	0.95	0.92	0.88	93%	0.87	0.96	0.29	0.84
Stimulant	0.96	0.95	0.90	92%	0.89	0.94	0.31	0.84
Sedative	0.88	0.79	0.69	88%	0.63	0.96	0.31	0.64
Hallucinogen	0.81	0.63	0.50	89%	0.42	0.97	0.42	0.46
Nicotine/Tobacco	0.92	0.88	0.84	88%	0.84	0.90	0.70	0.73

Model includes current CAT Dx and current severity score

MCA=Maximum classification accuracy, Se=sensitivity, Sp=specificity

For CAT-SUD-E stimulant = cocaine and amphetamine

2013). This level of inter-rater agreement is comparable to what we obtained for alcohol use disorder specifically (at the point of maximum classification current kappa = 0.36; lifetime kappa = 0.55). In general, agreement was higher for lifetime than current. Maximum classification accuracy was achieved at high levels of specificity and lower levels of sensitivity, which is not surprising, because these are relatively rare disorders.

The CAT-SUD-E functions well in comparison to other existing SUD screening tools. For instance, metrics for CAT overall SUD AUCs for current/lifetime (0.92/0.94) were higher than other reports across substance categories, including the Alcohol, Smoking and Substance Involvement Screening Test (ASSIST) use vs. abuse (0.84) and abuse vs. dependence (0.73) (Humeniuk et al., 2008) and the optimal Drug Abuse Screening Test (DAST) cut-off of 5/6 (0.93) (Gavin et al., 1989). The CAT substance-specific AUCs were on par with those reported for other tools for alcohol (CAT curr/life = 0.77/0.85, others = 0.70–0.98; ASSIST, AUDIT), cannabis (0.86/0.85, 0.62–0.96; ASSIST), opioids (0.90/0.95, 0.74–0.97; ASSIST), stimulants (0.91/0.96, 0.77–0.96; ASSIST), sedatives (0.77/0.88, 0.45–0.96; ASSIST), and hallucinogens (NA/0.88, NA) (Humeniuk et al., 2008; Moehring et al., 2019). In addition, CAT AUCs for nicotine (0.92/0.90) were relatively high, something not reported or assessed by other screening tools (ASSIST, AUDIT, DAST, TAPS) (Gavin et al., 1989; Humeniuk et al., 2008; McNeely et al., 2016; Moehring et al., 2019). In sum, CAT-SUD-E's psychometric properties suggest similar, or better, ability to detect SUDs than similar screening tools, noting that multiple screening/assessment tools, and greater than 4 min of administration time, would be needed to accomplish similar assessment outcomes. The CAT-SUD-E also has the clinical benefit of being integrated with the full CAT-MH suite to assess behavioral health currently, including providing a severity score for SUD and related mental health disorders and suicide risk that can be assessed over time (Gibbons and deGruy, 2019).

The CAT-SUD-E is designed to permit repeated evaluation of SUD severity over time with capacity for varying the reference time frame. Although the reproducibility of CAT-SUD-E scores was not determined, test-retest reliability of the CAT depression inventory (CAT-DI), which utilizes similar IRT-based methods, has been shown to exceed that of

the well-validated PHQ-9 depression screener (Beiser et al., 2016). Future work should address the utility of the CAT-SUD-E in monitoring changes in SUD diagnosis and severity, including in response to treatment. CAT-SUD-E scores also may be predictive of future development of SUDs in individuals who are currently at high risk of SUD. Prospective longitudinal studies are needed to address these questions. Additionally, given that SUDs often emerge in adolescence (National Institute on Drug Abuse, 2020; Schulenberg et al., 2017), there is a need to evaluate the diagnostic classification accuracy of the CAT-SUD-E in youth populations.

One limitation of the current study is the sample size ($n = 275$). Although comparable to the original CAT-SUD validation sample ($n = 297$; Gibbons et al., 2020) and enriched for SUD diagnoses, the low rates of certain specific SUDs (i.e., sedatives, hallucinogens) prevented examination of classification accuracy. While the current study addresses the most common and impairing SUDs observed in community and clinical settings (i.e., alcohol, cannabis, nicotine, opioids, stimulants), future studies in larger samples should attempt to address these gaps. Additionally, it is possible that selection of a fuller or different set of DSM-5-based items or different cut-points on the CAT-SUD-E would have yielded different patterns of results, though high correspondence with SCID results suggests the current set of items yielded strong clinical utility. Furthermore, research is needed on the utility and implementation factors associated with using the CAT-SUD-E in applied settings, including those where SUDs are frequently observed (e.g., criminal justice, community mental health, emergency departments). Finally, future studies should include greater representation of racial/ethnic minority participants – especially those identifying as Black or African American – than were enrolled here to test for equivalence of CAT-SUD-E predictive properties across racial groups.

5. Conclusions

We have developed a new approach for the diagnostic screening and measurement of SUD that builds on our earlier work in adaptive measurement of SUD. Our methodology synthesizes information from multiple related domains from mental health, trauma, and social support with

traditional SUD questions to provide a more comprehensive measure of SUD and adds to that current and lifetime overall and substance-specific SUD diagnoses. The CAT-SUD-E is highly predictive of a current and lifetime SUD diagnosis based on a structured clinical interview, accounting for substance-specific SUDs.

Declaration of Competing Interest

Dr. Gibbons has been an expert witness for the U.S. Department of Justice, Merck, Glaxo-Smith-Kline, Pfizer, and Wyeth and is a founder of Adaptive Testing Technologies, which distributes the CAT-MH™ battery of adaptive tests. The terms of this arrangement have been reviewed and approved by the University of Chicago in accordance with its conflict of interest policies. Dr. Hulvershorn has received support from an unrestricted grant to the National Network of Depression Treatment Centers from Greenwich Biosciences.

CRedit authorship contribution statement

Leslie A. Hulvershorn: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Zachary W. Adams:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Michael P. Smoker:** Data curation, Formal analysis, Investigation, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Matthew C. Aalsma:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing. **Robert D. Gibbons:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing.

Role of funding source

This work was supported by an administrative supplement from the National Institute of Mental Health (Washington DC) [Grant No. R01-MH100155] and a Substance Abuse and Mental Health Services Administration State Opioid Response grant to the Indiana Family and Social Services Administration [Grant No. TI081689-01]. In addition, this research was partially supported by the JCOIN cooperative, funded by the National Institute on Drug Abuse, National Institutes of Health. The authors gratefully acknowledge the collaborative contributions of NIDA and support from the following grant awards [NIDA UG1DA050070, U2CDA050098]. The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of NIDA or the participating sites.

References

American Psychiatric Association, 2013. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. American Psychiatric Association doi:10.1176/appi.books.9780890425596.

Authier, N., Balayssac, D., Sautereau, M., Zangarelli, A., Courty, P., Somogyi, A.A., Venat, B., Llorca, P.M., Eschaliere, A., 2009. Benzodiazepine dependence: focus on withdrawal syndrome. *Ann. Pharm. Fr.* 67 (6), 408–413.

Bayard, M., McIntyre, J., Hill, K., Woodside, J., 2004. Alcohol withdrawal syndrome. *Am. Fam. Physician* 69 (6), 1443–1450.

Beiser, D., Vu, M., Gibbons, R., 2016. Test-retest reliability of a computerized adaptive depression screener. *Psychiatr. Serv.* 67 (9), 1039–1041.

Brady, K.T., McCauley, J.L., Back, S.E., 2016. Prescription opioid misuse, abuse, and treatment in the United States: an update. *Am. J. Psychiatry* 173 (1), 18–26. doi:10.1176/appi.ajp.2015.15020262.

Creedon, T.B., Lê Cook, B., 2016. Access to mental health care increased but not for substance use, while disparities remain. *Health Aff.* 35 (6), 1017–1021.

First, M.B., Williams, J.B., Karg, R.S., Spitzer, R.L., 2015. *Structured Clinical Interview for DSM-5—Research Version (SCID-5 for DSM-5, Research Version; SCID-5-RV)*. American Psychiatric Association, pp. 1–94.

Gavin, D.R., Ross, H.E., Skinner, H.A., 1989. Diagnostic validity of the drug abuse screening test in the assessment of DSM-III drug disorders. *Br. J. Addict.* 84 (3), 301–307.

Gibbons, R.D., Alegria, M., Markle, S., Fuentes, L., Zhang, L., Carmona, R., Collazos, F., Wang, Y., Baca-García, E., 2020. Development of a computerized adaptive substance use disorder scale for screening and measurement: the CAT-SUD. *Addiction* 115 (7), 1382–1394. doi:10.1111/add.14938.

Gibbons, R.D., Bock, R.D., Hedeker, D., Weiss, D.J., Segawa, E., Bhaumik, D.K., Kupfer, D.J., Frank, E., Grochocinski, V.J., Stover, A., 2007. Full-information item bifactor analysis of graded response data. *Appl. Psychol. Meas.* 31 (1), 4–19. doi:10.1177/0146621606289485.

Gibbons, R.D., deGruy, F.V., 2019. Without wasting a word: Extreme improvements in efficiency and accuracy using computerized adaptive testing for mental health disorders (CAT-MH). *Curr. Psychiatry Rep.* 21 (8), 1–9.

Gibbons, R.D., Hedeker, D.R., 1992. Full-information item bi-factor analysis. *Psychometrika* 57 (3), 423–436.

Gibbons, R.D., Weiss, D.J., Pilkonis, P.A., Frank, E., Moore, T., Kim, J.B., Kupfer, D.J., 2012. Development of a computerized adaptive test for depression. *Arch. Gen. Psychiatry* 69 (11), 1104–1112.

Haffajee, R.L., Frank, R.G., 2018. Making the opioid public health emergency effective. *JAMA Psychiatry* 75 (8), 767–768.

Hasin, D.S., O'Brien, C.P., Auriacombe, M., et al., 2013. DSM-5 criteria for substance use disorders: recommendations and rationale. *Am. J. Psychiatry* 170 (8), 834–851.

Health Resources and Services Administration (HRSA). (2018). *State level projections of supply and demand for behavioral health occupations: 2016–2030*.

Hosmer, D.W., Lemeshow, S., 2000. *Applied Logistic Regression*. Wiley.

Humeniuk, R., Ali, R., Babor, T.F., Farrell, M., Formigoni, M.L., Jittiwutikarn, J., de Lacerda, R.B., Ling, W., Marsden, J., Monteiro, M., Nhwitiwa, S., Pal, H., Poznyak, V., Simon, S., 2008. Validation of the alcohol, smoking and substance involvement screening test (ASSIST). *Addiction* 103 (6), 1039–1047. doi:10.1111/j.1360-0443.2007.02114.x.

Katz, G., Lobel, T., Tetelbaum, A., Raskin, S., 2014. Cannabis withdrawal—a new diagnostic category in DSM-5. *Isr. J. Psychiatry* 51 (4), 270.

Kenny, P.J., Markou, A., 2001. Neurobiology of the nicotine withdrawal syndrome. *Pharmacol. Biochem. Behav.* 70 (4), 531–549.

Livne, O., Shmulewitz, D., Lev-Ran, S., Hasin, D.S., 2019. DSM-5 cannabis withdrawal syndrome: demographic and clinical correlates in US adults. *Drug Alcohol Depend.* 195, 170–177.

McNeely, J., Wu, L.T., Subramaniam, G., Sharma, G., Cathers, L.A., Svikis, D., Sleiter, L., Russell, L., Nordeck, C., Sharma, A., O'Grady, K.E., Bouk, L.B., Cushing, C., King, J., Wahle, A., Schwartz, R.P., 2016. Performance of the tobacco, alcohol, prescription medication, and other substance use (TAPS) tool for substance use screening in primary care patients. *Ann. Intern. Med.* 165 (10), 690–699. doi:10.7326/M16-0317.

Moehring, A., Rumpf, H.J., Hapke, U., Bischof, G., John, U., Meyer, C., 2019. Diagnostic performance of the alcohol use disorders identification test (AUDIT) in detecting DSM-5 alcohol use disorders in the general population. *Drug Alcohol Depend.* 204, 107530. doi:10.1016/j.drugalcdep.2019.06.032.

National Institute of Drug Abuse. (2020, June 2). *Principles of adolescent substance use disorder treatment*. <https://www.drugabuse.gov/publications/principles-adolescent-substance-use-disorder-treatment-research-based-guide/principles-adolescent-substance-use-disorder-treatment>

Pennay, A.E., Lee, N.K., 2011. Putting the call out for more research: the poor evidence base for treating methamphetamine withdrawal. *Drug Alcohol Rev.* 30 (2), 216–222.

Priester, M.A., Browne, T., Iachini, A., Clone, S., DeHart, D., Seay, K.D., 2016. Treatment access barriers and disparities among individuals with co-occurring mental health and substance use disorders: an integrative literature review. *J. Subst. Abuse Treat.* 61, 47–59.

Regier, D.A., Narrow, W.E., Clarke, D.E., Kraemer, H.C., Kuramoto, S.J., Kuhl, E.A., Kupfer, D.J., 2013. DSM-5 field trials in the United States and Canada, Part II: test-retest reliability of selected categorical diagnoses. *Am. J. Psychiatry* 170 (1), 59–70.

Schulenberg, J.E., Johnston, L.D., O'Malley, P.M., Bachman, J.G., Miech, R.A., Patrick, M.E., 2017. *Monitoring the Future National Survey Results on Drug Use, 1975–2016: Volume II, College Students and Adults Ages 19–55*. Ann Arbor: Institute for Social Research, The University of Michigan.

Substance Abuse and Mental Health Services Administration (US), 2016. Chapter 4, Early Intervention, Treatment, and Management of Substance Use Disorders Nov. US Department of Health and Human Services, Washington (DC) Available from: <https://www.ncbi.nlm.nih.gov/books/NBK424859/>.

University of Michigan Behavioral Health Workforce Research Center, 2018. *Estimating the Distribution of the U.S. Psychiatric Subspecialist Workforce*. UMSPH, Ann Arbor, MI.

Walker, R., Northrup, T.F., Tillitski, J., Bernstein, I., Greer, T.L., Trivedi, M.H., 2019. The stimulant selective severity assessment: a replication and exploratory extension of the cocaine selective severity assessment. *Subst. Use Misuse* 54 (3), 351–361.