

It is illegal to post this copyrighted PDF on any website.

Development of a Computerized Adaptive Test Suicide Scale—The CAT-SS

Robert D. Gibbons, PhD^{a,*}; David Kupfer, MD^b; Ellen Frank, PhD^b; Tara Moore, MA, MPH^c; David G. Beiser, MD, MS^d; and Edwin D. Boudreaux, PhD^e

ABSTRACT

Objective: Current suicide risk screening and measurement are inefficient, have limited measurement precision, and focus entirely on suicide-related items. For this study, a psychometric harmonization between related suicide, depression, and anxiety symptom domains that provides a more balanced and complete spectrum of suicidal symptomatology was developed. The objective of this article is to describe the results of the early stages of computerized adaptive testing development for a suicide scale and pave the way for the final stage of validation.

Methods: Data from psychiatric outpatients at the University of Pittsburgh and a community health clinic were collected from January 2010 through June 2012. 789 participants were enrolled in the calibration phase; 70% were female, and 30% were male. The rate of major depressive disorder as diagnosed by *DSM-5* was 47%. The item bank contained 1,008 items related to depression, anxiety, and mania, including 11 suicide items. Data were analyzed using a bifactor model to identify a core dimension between suicidal ideation, depression, anxiety, and mania items. A computerized adaptive test was developed via simulation from the actual complete item responses in 308 subjects.

Results: 111 items were identified that provided an extension of suicidality assessment to include statistically related responses from depression and anxiety domains that are syndromally associated with suicidality. All items had high loadings on the primary suicide dimension (average = 0.67; range, 0.49–0.88). Analyses revealed that a mean of 10 items (5–20) had a correlation of 0.96 with the 111-item scale, with a precision of 5 points on a 100-point scale metric. Preliminary validation data based on 290 clinician interviews revealed a 52-fold increase in the likelihood of current suicidal ideation across the range of the Computerized Adaptive Test Suicide Scale (CAT-SS).

Conclusions: The CAT-SS is able to accurately measure the latent suicide dimension with a mean of 10 items in approximately 2 minutes. Further validation against an independent clinician-administered assessment of suicide risk (ideation and attempts) and prediction of suicidal behavior is underway.

J Clin Psychiatry 2017;78(9):1376–1382

<https://doi.org/10.4088/JCP.16m10922>

© Copyright 2017 Physicians Postgraduate Press, Inc.

^aDepartments of Medicine and Public Health Sciences, The University of Chicago Biological Sciences, Illinois

^bWestern Psychiatric Institute and Clinic, The University of Pittsburgh, Pennsylvania

^cCenter for High Value Health Care, The University of Pittsburgh Medical Center, Pennsylvania

^dSection of Emergency Medicine, University of Chicago, Illinois

^eDepartments of Emergency Medicine, Psychiatry, and Quantitative Health Sciences, The University of Massachusetts Medical School, Worcester

*Corresponding author: Robert D. Gibbons, PhD, University of Chicago, 5841 S Maryland, Chicago, IL 60637 (rdg@uchicago.edu).

Broad-based screening and assessment of suicide risk within health care settings have been hampered by a dearth of reliable instruments that can be administered easily, quickly, and reliably. Even when rapid screening to identify non-negligible suicide risk is performed, health care professionals often are unfamiliar with how to further assess an individual who screens positive to derive a more precise measure of suicidal symptomatology.¹ This follow-up step of further assessing a positive frontline, case-identification screen often garners less attention in the literature but is as important as initial case identification. For example, item 9 of the Patient Health Questionnaire (PHQ-9)^{2,3} has been promoted as a potential frontline screener in medical settings. However, a positive screen simply identifies non-negligible suicide risk that requires additional follow-up; it does not provide much information on suicide risk severity or magnitude, and it does not assist with clinical decision-making other than to flag when further assessment is required.

Although health care professionals generally acknowledge that such follow-up assessment is needed, many are not prepared or trained to carry out this task, particularly those who are not trained mental health professionals.¹ A standardized paper-and-pencil measure such as the Beck Scale for Suicide Ideation (BSI)⁴ could be administered to help quantify severity, but the inconvenience of using, scoring, and interpreting such measures impedes widespread adoption. Resistance to using paper-based measures is further compounded by the movement away from paper-based medical records and to electronic health records (EHRs) and electronic patient-reported outcomes. Finally, another weakness of existing suicide measures reduces their utility: most, like the BSI, include only items related to suicidal ideation and suicidal behavior. This limits their ability to measure the full spectrum of suicidal symptomatology and requires the need for other psychiatric measures in order to be useful. For example, the most consistent risk factor for suicide remains having a psychiatric disorder, including unipolar depressive disorder,^{5,6} bipolar disorder,^{7,8} and anxiety disorders.^{5,9} An approach that blends measures of psychiatric symptoms, like depression, mania, and anxiety, with suicide-specific items can provide a more balanced and precise measure of risk across the spectrum from negligible to extreme suicide propensity. Indeed, there may well be symptoms of depression and anxiety that are key precursors to the development of suicidal thoughts, ideation, and behavior. Tests to measure suicidality can be likened to a general mathematical ability test with items ranging from simple arithmetic to advanced calculus. If only the calculus items are administered, the mathematical ability of senior level

- Suicide prevention is predicated on accurate risk detection and quantification.
- A single program that can be used for screening, quantification, and monitoring of risk and which enables more effective interventions while remaining feasible for use in both psychiatric and general medical settings would be truly transformative.
- The CAT-SS can provide precise measurement of suicidality based on self-reports in less than 2 minutes via the Internet and reproduces the results of structured clinical interviews so that identified patients can be further assessed by clinicians for their potential for self-harm.

college students majoring in mathematics or engineering may be accurately measured but the mathematical ability of elementary school students will not. Suicide measurement approaches that include only the upper range of suicide risk (ideation and behavior) are likely to be imprecise for those in the lower ranges. A precise quantification of suicide risk, therefore, requires assessing the full range of psychiatric symptoms that are related to the development of suicidal thoughts, ideation, and behavior.

Fortunately, many of the weaknesses that characterize current approaches to suicide screening and assessment can be addressed through computerized adaptive testing. Computerized adaptive testing allows for the creation of rapid, personally tailored screening and assessments that retain strong psychometric properties and, because they are administered electronically, decrease clinician and patient burden while at the same time accommodating the growing electronic transformation of health care.¹⁰ Computerized adaptive testing relies on Item Response Theory (IRT),¹¹ which models the relationship between a patient's responses to a series of items in terms of one or more latent variables that the test was designed to measure. Traditional mental health measurement based on classical test theory¹² fixes the items and allows the precision of measurement to vary from individual to individual. By contrast, IRT-based computerized adaptive testing fixes precision for different patients and for the same patient repeatedly measured over time and allows the items, both in number and content, to vary. Computerized adaptive testing adaptively selects a small set of items for each individual out of a much larger item bank, targeting precision by selecting successive items based on ability, trait, or impairment estimates derived from the responses to prior items administered. The net result is that we can both increase precision and decrease patient burden relative to traditional fixed-length tests.¹³

The National Institutes of Health has developed publicly available short-form tests and computerized adaptive tests (CATs) for physical, mental, and social well-being through an initiative called the Patient-Reported Outcomes Measurement Information System (PROMIS).¹⁴ These short-form tests and CATs have been constructed using IRT approaches. Some existing PROMIS CATs assess domains

relevant to suicide, such as depression and anxiety; however, a PROMIS CAT measuring suicidal symptomatology specifically is not currently available. The only published study¹⁵ of a CAT targeting suicidal ideation used the BSI⁴ with a sample of Dutch psychiatric patients. The 19 BSI suicidal ideation items could be reduced to an average of 4 items without losing discriminative ability.¹⁵ Although that article supports continued study of suicide-related CATs, the BSI is characterized by the aforementioned weakness of measuring only suicide-related items. In addition, the study was conducted in the Netherlands, so the performance of BSI CATs with individuals in the United States is unknown.

A fundamental problem with both the PROMIS CAT measures and the BSI CAT is their reliance on the assumption of unidimensionality of the underlying latent construct.¹⁶ Mental health constructs are inherently multidimensional, and applying unidimensional IRT to multidimensional data results in biased estimates of uncertainty, increased variability in severity estimates, and small item banks that dramatically minimize the possible benefits of computerized adaptive testing.¹⁶ To this end, we have developed a statistical methodology that extends computerized adaptive testing to multidimensional constructs.¹⁷ In particular, we use the bifactor model,^{16,18} which estimates a subject's location on the primary dimension of interest while permitting residual correlations among items within subdomains from which the items were drawn, such as depression as a primary domain with items drawn from subdomains of mood, somatization, cognitive impairment, and others. The net result is that we can obtain more realistic and unbiased estimates of uncertainty and develop large item banks (hundreds of items) that densely cover the entire continuum of the latent variable of interest. In this application, we provide an integration of suicide, depression, and anxiety symptoms that provides a unified and coherent primary dimension while also permitting disorder-specific subdomains within which items are allowed to be conditionally related.

Developing and validating the CAT Suicide Scale (CAT-SS) could propel efforts for suicide prevention by making available more efficient and precise quantification of risk, a goal that has been heralded by the National Action Alliance for Suicide Prevention as a top priority for the nation's research agenda.¹⁹ The process of creating such a CAT has several stages that include item bank development, calibration, CAT simulation, and validation against external criteria. The objective of this article is to describe the results of the early stages of CAT development for a suicide scale and pave the way for the final stage of validation.

METHODS

This study was conducted in compliance with the ethical principles of the Declaration of Helsinki, the US Food and Drug Administration guidelines, and the International Conference on Harmonization's Good Clinical Practices Guidelines. The Institutional Review Boards at both the

It is illegal to post this copyrighted PDF on any website.

University of Pittsburgh and the University of Chicago approved the study, and individuals signed a written informed consent form prior to initiation of study procedures.

Setting and Participants

Participants were male and female treatment-seeking outpatients between 18 and 80 years of age. Patients were recruited from 2 facilities, the Western Psychiatric Institute and Clinic at the University of Pittsburgh and a community clinic at DuBois Regional Medical Center. Psychiatric diagnoses were confirmed by medical records and the treating physician or clinician. Patients with and without a lifetime diagnosis of major depressive disorder (MDD) based on *DSM-5* criteria were included. Exclusion criteria included *DSM-5* schizophrenia, schizoaffective disorder, or psychosis; *DSM-5* organic neuropsychiatric syndromes (eg, Alzheimer's disease); *DSM-5* drug or alcohol dependence within the past 3 months (however, patients with episodic abuse related to mood episodes were not excluded); inpatient treatment status; and inability or unwillingness to provide informed consent. Complete details of the sample have been previously described.¹⁷

Item Bank

The item bank contained 1,008 items related to depression (452), mania (89), and anxiety (467), including 11 items measuring suicidal ideation. A key step in creating the original item bank¹⁷ was qualitative review of the items done by consensus among the members of the Pittsburgh research site. The items were selected based on a review of more than 100 existing depression or depression-related rating scales. Items were modified to refer to the previous 2-week period and to have consistent response categories. The majority of items were rated on a 5-point ordinal scale. Example items are provided in the online supplement of the previously published article.¹⁷

Data Collection

Data used to calibrate the bifactor model were collected from January 2010 through June 2012 as a part of the original study described by Gibbons and colleagues.¹⁷ They included item responses for a total of 789 subjects, 308 of whom had complete data for all 1,008 items. The remaining subjects took subsets of the items (252 items each) based on a balanced incomplete block design.²⁰

Validation Study

Preliminary validation data for the CAT-SS were obtained from the emergency departments of the University of Chicago (n = 155) and University of Massachusetts (n = 135). Columbia Suicide Severity Rating Scale (C-SSRS)²¹ clinical interviews and CAT-SS test administrations were conducted for all 290 patients.

Data Analytic Plan

The bifactor model,¹⁸ the first confirmatory multidimensional IRT model, was used for the primary

analysis and to build the CAT. It allows each item to measure the primary dimension (eg, suicide propensity) and a subdomain (eg, depression). This approach has computational and interpretational advantages over unrestricted exploratory item factor analytic models²² and extends CAT to the measurement of multidimensional constructs.¹⁶

First, the 11 suicidal ideation items were fitted to a unidimensional IRT model, and the distribution of the estimated scores was resolved into a mixture of normal distributions. The Bayesian Information Criterion (BIC) was used to select the number of component distributions, and the parameters were estimated by maximum likelihood. Second, each depression, mania, and anxiety item was used in a separate logistic regression to predict membership in the elevated suicidal component distribution. The top 100 items were included in the final item bank. A random forest²³ was used to assess the accuracy of the suicide prediction based on the depression, mania, and anxiety items alone. Third, a bifactor model was then fit to the 111 items (top 100 depression, mania, and anxiety items and the 11 suicidal ideation items) using subdomains of depression, mania, anxiety, and suicide. Fourth, based on the final bifactor model, a CAT was developed.¹⁷ The properties of the CAT were then determined by simulating CAT from the complete item-response data from the sample of 308 subjects. A finite mixture of normal distributions was then estimated from the final CAT-SS scores; the number of component distributions was selected based on minimizing the BIC.

For the validation component, we examined the association between the continuous CAT-SS score (underlying normal distribution spanning 6 points from -3 to 3) and the ordinal (5-point) C-SSRS ideation score and the C-SSRS lifetime suicide attempt rating using ordinal logistic regression. We also examined the relationship between the CAT-SS risk categories and the C-SSRS ideation score and C-SSRS lifetime suicide attempt rating using ordinal logistic regression. Finally, we examined the relationship between the CAT-SS risk categories and (a) any suicidal ideation, including passive or active (C-SSRS categories 1-5), (b) at least active ideation (C-SSRS categories 2-5), (c) C-SSRS suicide alert (C-SSRS categories 4-5 indicating plan or plan with intent), and (d) lifetime attempts. We computed rates, sensitivity and specificity, κ statistics, and 95% confidence intervals (CIs).

RESULTS

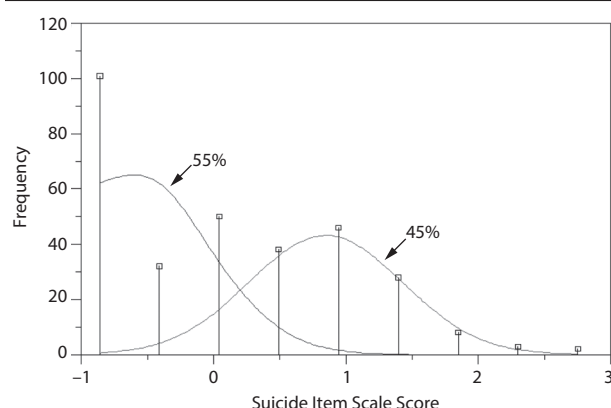
Descriptive Statistics

Subjects (N = 789) enrolled were 70% female and 30% male. The rate of MDD only was 27%; generalized anxiety disorder (GAD) only, 5%; other disorders (bipolar disorder, posttraumatic stress disorder, minor depression), 12%; and comorbid MDD and GAD, 18% (based on *DSM-5* criteria). Additional demographic characteristics of the sample have been previously reported.¹⁷

You are prohibited from making this PDF publicly available.

Table 1. The 11 Suicide Items From the Item Bank and Their Loading on the Latent Suicide Dimension

Item Description	Factor Loading
1. Have you felt that life was not worth living?	0.94
2. Did you want to be dead at times?	0.93
3. How often did you feel that others would be better off if you were dead?	0.91
4. Did you think that life was not worth living?	0.90
5. Have you found yourself wishing you were dead and away from it all?	0.90
6. I felt that others would be better off if I were dead.	0.89
7. I felt that life was empty or wondered if it was worth living.	0.87
8. Have you felt that life wasn't worth living?	0.87
9. Did you think about taking your own life?	0.77
10. I had a reason for living.	0.63
11. Did you think it was wonderful to be alive?	0.59

Figure 1. Finite Mixture Distribution for Estimated Scores on the 11-Item Suicide Scale^a

^aA mixture of 2 normal distributions with different means (normal and elevated), pooled standard deviation, and different mixing proportions. The high suicidal ideation component distribution has an elevated mean containing 45% of the sample. The smooth curves are the 2 estimated normal distributions, and the vertical lines represent the histogram (observed frequency distribution) of the estimated scores.

Characterization of the Suicide Dimension

The 11 suicide items alone formed a single unidimensional construct with loadings ranging from 0.59 to 0.94 with a mean loading of 0.84 (see Table 1). While all items had strong loadings on the underlying suicide dimension, the 2 positively worded items had lower discrimination than the negatively worded items.

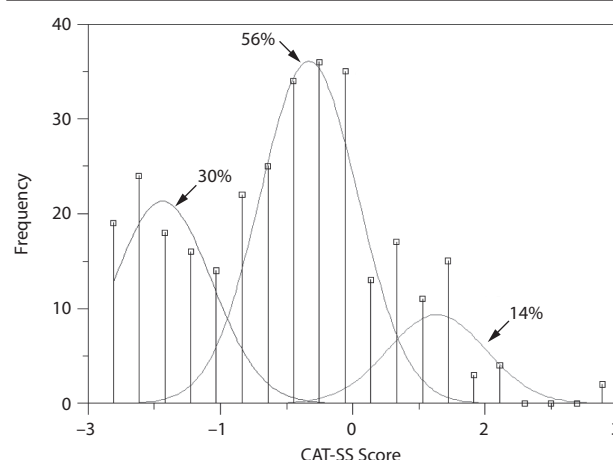
The distribution of the estimated scores resolved into a mixture of 2 normal component distributions (see Figure 1), with significant improvement in fit over a single normal distribution ($\chi^2_2 = 26.08$, $P < .0001$). Forty-five percent of the sample was in the elevated component distribution.

Creation of Expanded Suicide Item-Bank

The 997 items (1,008 original items minus 11 suicide items) were each tested for association with the elevated suicidal component distribution in Figure 1 using logistic regression. Odds ratios (ORs) for the top 100 items ranged from 2.4 to 42.0. Example items and their corresponding ORs are displayed in Table 2. These 100 items were made

Table 2. Example Depression Items and Their Association With Elevated Suicidal Severity

Item Description	Odds Ratio
How much were you distressed by feeling that people were unfriendly or disliked you?	2.5
How much did feelings of sadness or depression interfere with your everyday functioning?	5.0
Did you feel inadequate in everyday relationships, even with people like coworkers, sales people, or peers?	10.0
Did you see the future as very bleak?	15.0
Did you have the feeling that you just didn't have what it takes anymore?	20.0
Did you feel you wanted to give up trying?	40.0

Figure 2. Finite Mixture Distribution for Final CAT-SS Scores^a

^aA mixture of 3 normal distributions with different means (none, mild, and elevated), pooled standard deviation, and different mixing proportions. The high suicidal propensity component distribution has an elevated mean containing 14% of the sample, 56% are in the intermediate group (ie, mild ideation or depression and anxiety precursors), and 30% had no evidence of suicidal thoughts or propensity. The smooth curves are the 3 estimated normal distributions, and the vertical lines represent the histogram (observed frequency distribution) of the estimated scores. Abbreviation: CAT-SS = Computerized Adaptive Test Suicide Scale.

up of depression and anxiety items only; no mania items were retained in the reduced item bank. This is not to say that mania is not related to suicidal ideation, but rather the smaller set of mania items (89 of 1,008) were not among the subset of 100 items that were most highly related to suicidal ideation. A multivariate analysis based on a random forest revealed that these depression and anxiety symptoms predicted elevated suicidal symptoms with cross-validated sensitivity of 0.81 and specificity of 0.90.

Calibration

The fit of the bifactor model was significantly improved over the unidimensional alternative ($\chi^2_{111} = 2,161$, $P < .0001$). All 111 items had strong loadings on the primary suicide dimension (mean = 0.67; range, 0.49–0.88), indicating that the primary dimension provided a core dimension characterized by a synthesis of strongly related depression, anxiety, and suicide symptoms.

It is illegal to post this copyrighted PDF on any website.

Simulated CAT

Using the data for the 308 subjects with complete data, simulated CAT (ie, simulating CAT administration from the actual complete item responses) revealed that a mean of 10 items (range, 5–20) provided a correlation of 0.96 with the 111-item scale total score (from the complete test administrations) with precision of 5 points on a 100-point scale metric.

Empirical Distribution of the CAT-SS Scores

The distribution of the CAT-SS scores based on the sample of 308 subjects with complete data is displayed in Figure 2. Figure 2 indicates the presence of 3 component distributions: low or no risk (30%), possible or intermediate risk (56%), and high risk (14%). Thresholds correspond to scores of 34 (low vs intermediate risk) and 71 (intermediate vs high risk) in the 100-point metric. The mixture of 3 normal distributions improved the fit over both a single normal distribution ($\chi^2_2 = 21.90$, $P < .0001$) and a mixture of 2 normal ($\chi^2_2 = 18.04$, $P < .0001$) distributions.

Example CAT-SS Administrations

Supplementary eTable 1 (available at PSYCHIATRIST.COM) presents examples of 3 CAT-SS sessions for patients with high, moderate, and low suicidal severity. All sessions begin with the item “Have you felt that life was not worth living?” and adapt from there. Doing so ensures that there is always at least 1 suicide item in each adaptive testing session. In both cases, the CAT terminated when the uncertainty was at or below 5 points on the 100-point transformed scale.

Domains and Thresholds

Supplementary eTable 2 presents the domains, subdomains, facets, and mean thresholds for positive symptomatology, with higher thresholds indicating association with higher suicidal severity. The suicidal ideation items have the highest thresholds, indicating that they represent the most severe items in the scale, followed by items related to helplessness, guilt, and somatic anxiety and behavior. Both depression items and anxiety items are associated with high levels of suicidal severity. On the low end of the scale are items related to interpersonal behavior, cognitive information deficits, low activity, and negative affect (see Supplementary eTable 2).

Validation Study

In this emergency department population, 168 subjects were categorized as no risk (58%), 90 subjects as low risk (31%), and 32 as high risk (11%) using the CAT-SS thresholds based on the mixture distribution. Based on the C-SSRS interviews, 44 subjects (15%) had any ideation, 26 (9%) had active ideation, and 16 (6%) generated a suicide alert (plan or plan with intent). Forty-one subjects (14%) had a lifetime attempt (there were too few current suicide attempts to provide a meaningful analysis). In this sample, the CAT-SS took a mean of 110 seconds with a median of 11 items (range, 5–19, mean = 11.24).

A unit increase in CAT-SS score had an OR of 8.61 (95% CI, 5.15–14.38, $P < .0001$) for a category increase on the 5-category ordinal C-SSRS ideation scale, or a 52-fold increase across the entire CAT-SS scale for a category increase in C-SSRS ideation category. A unit increase in CAT-SS score had an OR of 2.28 (95% CI, 1.70–3.06, $P < .0001$) for a lifetime attempt, or a 14-fold increase across the entire CAT-SS scale. In terms of CAT-SS risk categories (none, low, high), the OR was 16.74 (95% CI, 8.52–32.87, $P < .0001$) per risk category for the C-SSRS ideation ordinal score, or a 33-fold increase from no risk to high risk. For lifetime attempts, the OR was 3.32 (95% CI, 2.08–5.29, $P < .0001$), or a 7-fold increase from no risk to high risk.

For any ideation on the C-SSRS, rates were 0.0% (0/168) for the CAT-SS no-risk, 23.3% (21/90) for the CAT-SS low-risk, and 71.9% (23/32) for the CAT-SS high risk group. Contrasting the no-risk and high-risk groups on the CAT-SS had a sensitivity of 1.00 (95% CI, 0.85–1.00) and a specificity of 0.95 (95% CI, 0.93–0.95) for the C-SSRS any-ideation categorization with agreement of $\kappa = 0.81$ (95% CI, 0.66–0.81). For active ideation on the C-SSRS, rates were 0.0% (0/168) for the CAT-SS no-risk, 10.0% (9/90) for the CAT-SS low-risk, and 53.1% (17/32) for the CAT-SS high-risk group. Contrasting the no-risk and high-risk CAT-SS groups had a sensitivity of 1.00 (95% CI, 0.79–1.00) and a specificity of 0.92 (95% CI, 0.90–0.92) for the C-SSRS active ideation categorization with agreement of $\kappa = 0.66$ (95% CI, 0.49–0.66). For the C-SSRS suicide alert (plan or plan and intent), rates were 0.0% (0/168) for the CAT-SS no-risk, 5.6% (5/90) for the CAT-SS low-risk, and 34.4% (11/32) for the CAT-SS high-risk group. Contrasting the no-risk and high-risk CAT-SS groups had a sensitivity of 1.00 (95% CI, 0.70–1.00) and a specificity of 0.89 (95% CI, 0.87–0.89) for the C-SSRS warning with agreement of $\kappa = 0.47$ (95% CI, 0.30–0.47). For the C-SSRS lifetime attempt rating, rates were 4.8% (8/168) for the CAT-SS no-risk, 24.4% (22/90) for the CAT-SS low-risk, and 34.4% (11/32) for the CAT-SS high-risk group. Contrasting the no risk and high CAT-SS risk groups had a sensitivity of 0.58 (95% CI, 0.36–0.78) and a specificity of 0.88 (95% CI, 0.86–0.91) for lifetime attempts with agreement of $\kappa = 0.35$ (95% CI, 0.17–0.52).

DISCUSSION

The CAT-SS is able to accurately measure suicidal severity with a median of 11 items in less than 2 minutes. The validation study revealed that the CAT-SS accurately tracks suicidal ideation across the severity range of clinician-rated C-SSRS categories. Agreement is highest with clinical ratings of any ideation, but there is still a strong association between the CAT-SS and active ideation, ideation with plan or plan and intent, and even lifetime suicide attempts. As such, it can be used to reliably assess suicide risk both in clinic environments and through remote monitoring without clinician burden and minor patient burden.

The next stage in the scale's development is to administer it with a larger sample of heterogeneous patients (eg, patients

coming to an emergency department) and to validate it against independent criteria, including independent clinician-administered suicide risk assessment for both ideation and current suicidal behavior and the prediction of prospective suicidal behavior within 6 months. These studies should also assess test-retest reliability and sensitivity to change over time. Previous study of test-retest reliability of the CAT depression test revealed higher test-retest reliability ($r=0.92$) relative to traditional fixed-length tests (PHQ-9 $r=0.84$) despite the use of different items upon repeat testing.²⁴ The strength of the CAT-SS is the inclusion of depression and anxiety symptoms, which map onto the primary and suicidal ideation dimension, so that a more complete view of suicidal risk can be obtained. As such, it may provide an early warning system for patients at risk for becoming suicidal prior to the emergence of the report of suicidal ideation or behavior.

It is important to note that not all depression, mania, and anxiety symptoms are related to suicide. In fact, none of the mania symptoms were retained in the final item bank, and only a small subset of depression and anxiety symptoms were retained. An advantage of computerized adaptive testing is that new symptoms can be added to the bank, calibrated, and then added to the CAT-SS once sufficient data are available. Further study of the CAT-SS in patients with bipolar disorder may reveal that a subset of mania symptoms are predictive of suicide symptoms and can be added to further optimize the measurement of suicide risk in patients with bipolar disorder.

Other advantages of computerized adaptive testing include the ability to repeatedly administer the test within short durations without risk of response bias produced by the repeated administration of the same items. We can also determine whether the CAT-SS is valid in different populations using differential item functioning,^{25,26} which can be done for administration in different cultures in different languages and also for different indications such as postpartum depression for which certain symptoms may be confounded by physical symptoms of pregnancy. Unlike

traditional tests, which require repeated administration of the same items regardless of results of previous test results, computerized adaptive testing allows us to begin the next administration using the previous CAT-SS score as a starting point. This characteristic will further improve the efficiency of the CAT-SS and lead to even further reduction in patient burden. Finally, the ability to administer the CAT-SS via the Internet using a cloud computing platform further decreases barriers to testing. Where the ability to provide a timely response is available, remote screening of suicide risk is viable. More generally, the addition of the CAT-SS to the Computerized Adaptive Testing Mental Health (CAT-MHTM), which includes adaptive tests for depression, anxiety, and mania, will dramatically improve mental health screening and measurement in real-world settings in less than 8 minutes.

This study has several limitations. First, we have yet to show predictive validity based on suicidal behavior in the 3 to 6 months following the index episode assessment. Second, the CAT-SS needs to be augmented with other suicide risk predictors such as demographic variables and previous suicidal behavior in order to provide a suicide risk prediction system. The CAT-SS is just one component of this risk prediction system. Third, the validity of the CAT-SS in different languages, cultures, and settings (eg, perinatal) has yet to be demonstrated.

CONCLUSIONS

We have developed a new approach to the dimensional measurement of suicidal severity that may serve as an important element in the prediction of suicidal risk. Our methodology synthesizes information on depression, anxiety, and suicide symptomatology to provide a dimensional measure of suicidal severity that may be predictive of future suicidal behavior even before the emergence of suicidal thoughts and ideation. The next step in this research is to administer the scale in heterogeneous populations and relate it to the prediction of future suicidal behavior.

Submitted: May 7, 2016; accepted September 1, 2016.

Online first: May 9, 2017.

Potential conflicts of interest: Dr Frank has received royalties or honoraria from the American Psychological Association and Guilford Press, has served on an advisory board for Servier International, and has financial interests in Adaptive Testing Technologies (www.adaptivetestingtechnologies.com), through which the CAT-SS test will be made available, and in HealthRhythms. Dr Kupfer holds joint ownership of copyright for the Pittsburgh Sleep Quality Index, has received an honorarium from and served on an advisory board for Servier International, and is a stockholder in Minerva Neuroscience, AliphCom, Adaptive Testing Technologies, and Health-Rhythms. Dr Gibbons has been an expert witness for Merck, Pfizer, GlaxoSmithKline, the US Department of Justice, and Wyeth and has financial interests in Adaptive Testing Technologies, which distributes the CAT-MHTM battery of adaptive tests. Dr Boudreaux has financial interests in Polaris

Health Directions (which was not associated with this study). Dr Beiser and Ms Moore have no conflicts of interest to declare.

Funding/support: This study was funded through a grant from the National Institute of Mental Health (Washington DC) R01-MH66302. The NIMH funded the collection of the data and the development of the original computerized adaptive testing methodology.

Role of the sponsor: The sponsor had no role in design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript.

Previous presentation: The data were presented at the Plenary Session of the International Summit on Suicide Research; October 11, 2015; New York, New York.

Acknowledgments: The authors acknowledge the editorial assistance of Brianna L. Haskins, MS, of the Department of Emergency Medicine, The University of Massachusetts Medical School, Worcester (compensated for

effort). Ms Haskins has no conflicts of interest to declare.

Supplementary material: Available at PSYCHIATRIST.COM.

REFERENCES

1. Betz ME, Arias SA, Miller M, et al. Change in emergency department providers' beliefs and practices after use of new protocols for suicidal patients. *Psychiatr Serv*. 2015;66(6):625–631.
2. Kocalevent RD, Hinz A, Brähler E. Standardization of the depression screener Patient Health Questionnaire (PHQ-9) in the general population. *Gen Hosp Psychiatry*. 2013;35(5):551–555.
3. Bauer AM, Chan YF, Huang H, et al. Characteristics, management, and depression outcomes of primary care patients who endorse thoughts of death or suicide on the PHQ-9. *J Gen Intern Med*. 2013;28(3):363–369.
4. Beck AT, Steer RA. *BSI, Beck Scale for Suicide Ideation: Manual*. San Antonio, TX: Pearson; 1991.

It is illegal to post this copyrighted PDF on any website.

5. Davidson CL, Wingate LRR, Grant DM, et al. Interpersonal suicide risk and ideation: the influence of depression and social anxiety. *J Soc Clin Psychol*. 2011;30(8):842–855.
6. Hawton K, Casañas I Comabella C, Haw C, et al. Risk factors for suicide in individuals with depression: a systematic review. *J Affect Disord*. 2013;147(1–3):17–28.
7. Cassidy F. Risk factors of attempted suicide in bipolar disorder. *Suicide Life Threat Behav*. 2011;41(1):6–11.
8. McIntyre RS, Muzina DJ, Kemp DE, et al. Bipolar disorder and suicide: research synthesis and clinical translation. *Curr Psychiatry Rep*. 2008;10(1):66–72.
9. Bolton JM, Cox BJ, Afifi TO, et al. Anxiety disorders and risk for suicide attempts: findings from the Baltimore Epidemiologic Catchment area follow-up study. *Depress Anxiety*. 2008;25(6):477–481.
10. Weiss DJ. Adaptive testing by computer. *J Consult Clin Psychol*. 1985;53(6):774–789.
11. Lord FM, Novick MR. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley Publishing Company; 1968.
12. Traub R. Classical test theory in historical perspective. *Educational Measures: Issues and Practice*. 1997;16(4):8–14.
13. Meijer RR, Nering ML. Computerized adaptive testing: overview and introduction. *Appl Psychol Meas*. 1999;23(3):187–194.
14. PROMIS instrument development and psychometric evaluation scientific standards. HealthMeasures website. http://www.nihpromis.org/Documents/PROMIS_Standards_050212.pdf. Accessed May 09, 2015.
15. De Beurs DP, de Vries ALM, de Groot MH, et al. Applying computer adaptive testing to optimize online assessment of suicidal behavior: a simulation study. *J Med Internet Res*. 2014;16(9):e207.
16. Gibbons RD, Bock D, Hedeker D, et al. Full-information item bifactor analysis of graded response data. *Appl Psychol Meas*. 2007;31(1):4–19.
17. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry*. 2012;69(11):1104–1112.
18. Gibbons RD, Hedeker D. Full information item bi-factor analysis. *Psychometrika*. 1992;57(3):423–436.
19. Data and Surveillance Task Force of the National Action Alliance for Suicide Prevention. Improving national data systems for surveillance of suicide-related events. *Am J Prev Med*. 2014;47(suppl 2):S122–S129.
20. Cochran WG, Cox GM. *Experimental Designs*. New York, NY: Wiley; 1957.
21. Posner K, Brown GK, Stanley B, et al. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am J Psychiatry*. 2011;168(12):1266–1277.
22. Bock RD, Aitkin M. Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*. 1981;46(4):443–459.
23. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5–32.
24. Beiser D, Vu M, Gibbons R. Test-retest reliability of a computerized adaptive depression screener. *Psychiatr Serv*. 2016;67(9):1039–1041.
25. Thissen D, Steinberg L, Wainer H. Detection of differential item functioning using the parameters of item response models. In: Holland P, Wainer H, eds. *Differential Item Functioning*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc; 1993:67–100.
26. Cai L, Yang JS, Hansen M. Generalized full-information item bifactor analysis. *Psychol Methods*. 2011;16(3):221–248.

See supplementary material for this article at PSYCHIATRIST.COM.

You are prohibited from making this PDF publicly available.



Supplementary Material

Article Title: Development of a Computerized Adaptive Suicide Scale—the CAT-SS

Author(s): Robert D. Gibbons, PhD^{a,*}; David Kupfer, MD^b; Ellen Frank, PhD^b; Tara Moore, MA, MPH^c; David G. Beiser, MD, MS^d; and Edwin D. Boudreaux, PhD^e

DOI Number: <https://doi.org/10.4088/JCP.16m10922>

List of Supplementary Material for the article

1. [eTable 1](#) Example Testing Sessions for Symptoms in the Past 2 Weeks
2. [eTable 2](#) Domains, Subdomains, Facets, and Average thresholds for Positive Symptomatology

Disclaimer

This Supplementary Material has been provided by the author(s) as an enhancement to the published article. It has been approved by peer review; however, it has undergone neither editing nor formatting by in-house editorial staff. The material is presented in the manner supplied by the author.

Supplementary eTable 1: Example Testing Sessions for Symptoms in the Past 2 Weeks

High Suicidal Severity	
Have you felt that life was not worth living?	Quite a bit
How much were you distressed by feelings of worthlessness?	Extremely
How often did you feel that others would be better off if you were dead?	Often
How often did you feel sad?	All of the time
How much have you felt nothing was interesting or fun?	Quite a bit
How much were you distressed by feeling blue?	Extremely
How much were you distressed by feeling hopeless about the future?	Quite a bit
How much have you felt like a failure?	Extremely
Has feeling depressed interfered with what you usually do?	Most of the time
I felt gloomy.	Extremely
How much have you felt withdrawn from others?	Quite a bit
I felt that I could not shake off the blues even with help from my family or friends.	Very much
I was unhappy.	Always
How much have you felt confused?	Quite a bit
Score=89.9, Precision 4.9, High Risk	
Moderate Suicidal Severity	
Have you felt that life was not worth living?	Moderately
I felt depressed.	Some of the time
Did you experience long periods of sadness?	Sometimes
Has feeling depressed interfered with what you usually do?	Some of the time
I felt that I could not shake off the blues even with help from my family or friends.	Somewhat
How much of the time have you felt so down in the dumps that nothing could cheer you up?	Some of the time
Did you feel isolated from others?	Some of the time
How much difficulty have you been having in the area of isolation or feelings of loneliness?	Moderate
How much were you distressed by the idea that something was wrong with your mind?	Moderately
How much have you been disappointed in yourself?	Moderately
My life lacked meaning and purpose.	Somewhat
Score 63.5, Precision 5.0., Moderate Risk	

Supplementary eTable 1: Example Testing Sessions for Symptoms in the Past 2 Weeks (continued).

Low Suicidal Severity	
Have you felt that life was not worth living?	Not at all
How sad did you feel?	A little bit
How much were you distressed by repeated unpleasant thoughts?	A little bit
How much were you distressed by feelings of guilt?	Not at all
I was unhappy.	Occasionally
Did you feel isolated from others?	A little of the time
How much have you felt that nothing was enjoyable?	Not at all
How much of the time did you have difficulty reasoning and solving problems; for example, making plans, making decisions, learning new things?	A little of the time
I worried that my condition will get worse.	A little bit
Have you been feeling out of sorts?	No more than usual
I felt that life was empty or wondered if it was worth living.	Not at all
You felt you had no purpose, as if everything had lost its significance?	No
Have you had any reason to wonder if you were losing your mind?	No
Severity 35.9, Precision 5.0, Low Risk	

**Supplementary eTable 2: Domains, Subdomains, Facets, and Average thresholds
for Positive Symptomatology***

1.	Depression\Behavior\Interpersonal	(2 items, average threshold = -1.06)
2.	Depression\Cognition\Information Deficits	(3 items, average threshold = -1.01)
3.	Depression\Behavior\Low Activity	(2 items, average threshold = -1.01)
4.	Depression\Mood\Negative Affect	(17 items, average threshold = -0.99)
5.	Anxiety\Mood\Fear & Anxiety	(7 items, average threshold = -0.83)
6.	Depression\Cognition\Social Cognition	(8 items, average threshold = -0.83)
7.	Depression\Cognition\Impaired View	(9 items, average threshold = -0.76)
8.	Depression\Behavior\Low Energy	(2 items, average threshold = -0.70)
9.	Depression\Cognition\Hopelessness	(6 items, average threshold = -0.67)
10.	Depression\Mood\Positive Affect	(10 items, average threshold = -0.54)
11.	Anxiety\Mood\Anxious Misery	(1 item, average threshold = -0.52)
12.	Anxiety\Cognition	(18 items, average threshold = -0.41)
13.	Depression\Treatment\Perceived Need	(1 item, average threshold = -0.39)
14.	Depression\Cognition\Helplessness	(5 items, average threshold = -0.36)
15.	Depression\Cognition\Guilt	(3 items, average threshold = -0.31)
16.	Anxiety\Somatic\Specific Symptoms	(2 items, average threshold = -0.30)
17.	Anxiety\Behaviors	(5 items, average threshold = -0.27)
18.	Depression\Cognition\Information Unproductive	(1 item, average threshold = -0.06)
19.	Depression\Suicidality\Suicidal Ideation	(11 items, average threshold = 0.11)
*The higher the threshold, the more severe the suicidal content of the symptom.		