

INTERPRETATION OF CAT-MH[®] TEST SCORES

GUIDELINES FOR A GENERAL POPULATION

CONTENTS

The Value of the CAT-MH®	3
Understanding CAT-MH® Severity and Precision.....	3
Table 1: How to Use Precision Measures to Interpret Change.....	4
Modules	5
Suicidality	5
Brief Suicidality Screener: C-SSRS.....	5
Suicide Scale: CAT-SS	5
Depression	7
Brief Diagnostic Screener for Major Depressive Disorder: CAD-MDD	7
Depression Inventory: CAT-DI	7
The PHQ-9.....	8
Anxiety	9
CAT-ANX.....	9
Mania/Hypomania	10
CAT-MANIA	10
Substance Use Disorder	11
CAT-SUD For Longitudinal Use	11
CAT-SUD-E For DIAGNOSTIC Screening.....	12
Post-Traumatic Stress Disorder	13
CAT-PTSD FOR LONGTUDINAL USE [RESTRICTED, REPLACED WITH PTSD-E ON 12/19/2021]	13
PTSD-DX FOR SCREENING	13
PTSD-E FOR SEVERITY MEASUREMENT AND LONGITUDINAL ASSESSMENTS	13
Psychosis	14
CAT-PSYCHOSIS	14
Adult ADHD	15
CAT-ADHD	15
Social Determinants of Health.....	16
CAT-SDoH.....	16

These recommended guidelines are based on client data and observations. To obtain the benefit from using the CAT-MH®, your organization should determine whether these interpretation guidelines require modifications to meet the needs and goals of your organization’s efforts and user populations. For additional information, visit www.adaptivetestingtechnologies.com, email info@adaptivetestingtechnologies.com, or call (312) 878-6490.

The CAT-MH® is a uniquely effective and accurate approach to measuring, tracking, and screening for a variety of conditions and other mental health indicators including:

- Depression, including Major Depressive Disorder
- Anxiety
- Mania/hypomania
- Substance Use Disorder
- Post-Traumatic Stress Disorder
- Psychosis
- Adult ADHD
- Social Determinants of Health
- Suicidality

The full battery of modules can be administered to a client in 10 to 15 minutes on any web-enabled device via a HIPAA compliant, secure Amazon Web Services (AWS) platform. However, many users target specific modules to their population, resulting in a comprehensive assessment completed in even short time periods.

Because the CAT-MH® has low client burden, it is ideal for tracking symptoms remotely between visits, over time, and ultimately, screening clients to determine their level of need and type of care. However, the real power of the CAT-MH® is the accuracy and clarity that it brings to assessment.

The CAT-MH® has been validated against structured clinical interviews with precision matching or exceeding the agreement between two trained clinician interviewers. This means that the results of the CAT-MH® are aligned with what a trained clinician would find when completing a full structured clinical diagnostic interview, taking an hour or more.

UNDERSTANDING CAT-MH® SEVERITY AND PRECISION

Unlike most traditional measures, the CAT-MH® results include three important metrics:

- **Severity:** a numeric score that indicates how severe a client's presentation for a condition is on a 100-point scale where 0 represents the lowest level of severity and 100 represents the highest level of severity
- **Category:** a categorization of a client's severity score into the categories: normal, mild, moderate, or severe or low risk, medium risk, high risk, depending on the module
- **Precision:** a number that represents the degree of uncertainty of the severity score. Adaptive testing allows the precision to be fixed to 5 points on a 100-point scale, where 0 would represent absolute certainty and 100 would represent absolute uncertainty

The value of these metrics lies in their ability to help a clinician understand where a client is at in a precise way. This precision has important implications in practice because it allows you to measure change over time in a statistically meaningful way, at the individual patient level.

The severity score and accompanying category tell the clinician how a given client is presenting *in relation to* the general population that a given module has been tested on. Clinicians can use this information in consultation with the client to

triage clients into higher or lower levels of care and determine, when multiple conditions are present, which conditions should be prioritized.

The precision is useful in that it gives clinicians the ability to understand how much certainty they should hold about a given severity score. This is valuable for a few reasons. First, a clinician who sees a severity score of 75 with a precision score of 5.0, can know that – scores in the range of 70-80 are within the limits of precision for this measured score of 75. If we were to repeat this assessment 100 times, the true severity would be contained with-in the interval of 65 to 85, 95% of the time (i.e., 95% confidence interval = severity score plus/minus 1.965*precision). Second, the precision can be used to determine if a individual’s change over time is clinically and / or statistically significant.

When using the CAT-MH®, there are three ways of categorizing change. It can be: uncertain, clinically significant, and statistically significant. Table 1 below provides a useful example with definitions of each type of change.

TABLE 1: HOW TO USE PRECISION MEASURES TO INTERPRET CHANGE

Case	Precision	Test 1	Test 2	Absolute Distance	Type of Change	Definition of Change Type
A	5.0	75	70	$ 75-70 = 5$	Uncertain	A change is not clinically or statistically significant when the second score is still within the bounds of uncertainty (5 points). In this case, any second score between 80 and 70 (inclusive) would be considered an insignificant change.
B	5.0	75	69	$ 75-69 = 6$	Clinically Significant	A change is clinically significant when the absolute distance between the first and second scores is larger than the precision score. Change at this level may be clinically meaningful, but it hasn't met the higher bar of statistical significance. In this case, any second score below 70 or above 80 would be <i>at least</i> clinically significant.
C	5.0	75	64	$ 75-64 = 11$	Statistically Significant	A change is statistically significant when the absolute difference between the first and second scores is more than 2x the level of precision. Where the level of precision is 5, a statistically significant change would need to be more than 10 points away from the first score. Change at this level is likely to present real and meaningful change. In this case, any second score below 65 or above 85 would be statistically significant.

With this information, clinicians have the opportunity to truly understand how an individual is doing on a monthly, weekly, or even daily basis, and tailor care accordingly. Longitudinal analysis of CAT-MH® scores, when used in treatment planning, can help a clinician understand the efficacy of modalities and strategies being used. Traditional mental health measurement scales do not provide uncertainty estimates and therefore cannot be used to precisely estimate change at the individual patient level.

The CAT-MH® includes two modules related to suicidality - the C-SSRS brief screener and the CAT-SS. These two modules provide different types of data, and because of this, are particularly useful when used in conjunction. Both the C-SSRS and the CAT-SS may trigger a suicide warning but under different conditions (discussed below).

BRIEF SUICIDALITY SCREENER: C-SSRS

The C-SSRS Suicide Screen is a 2 to 4 item screening measure developed in conjunction with the Department of Psychiatry at Columbia University. It can include questions pertaining to suicidal thought, intent, plan, and behavior. A positive screen has been shown to result in a 9-fold increase in the likelihood of a suicide attempt in the following 6 months. This tool is best used to assess for imminent risk of suicide, within the next week.

Metrics Provided: *Suicide alert (positive or negative)*

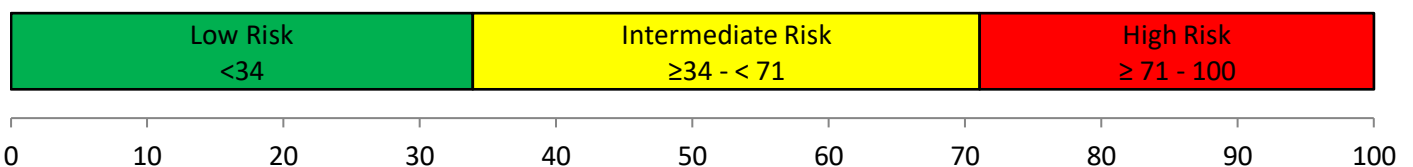
Notes:

1. This tool will trigger a suicide warning when a client has specific intentions, plans, or behaviors (past, present, or future) related to suicidality.
2. This tool will first ask about suicidal ideation and ask up to 4 questions to generate a positive or negative screen. The C-SSRS relies on the endorsement of 2 categories (ideation, intent, plan, behavior) unless the individual responds in the negative for ideation but in the positive for behavior.
3. Note that a patient may have very severe suicidal thoughts and ideation, but if there is no intent, plan, or behavior, the test will result in the same negative result as a patient without any suicidal thoughts or ideation. This is a serious limitation of traditional suicide measures that do not provide a crosswalk to other symptoms that are precursors of suicidality and passive suicide items resulting in a continuous severity score and risk categories like the CAT-SS.

SUICIDE SCALE: CAT-SS

The CAT-SS is a dimensional severity measure that requires, on average, 10 questions to effectively assess for suicidality. This scale provides a crosswalk between depression and anxiety precursors to suicidality and suicidal ideation. Together, questions from different domains provide a holistic measure of suicidality. This tool is best used to measure suicidality, stratify risk of a future suicide attempt over the next 3 to 6 months, and assess change in the severity of suicidality over time.

Severity Score Thresholds



Metrics: *Severity, Category, Precision, Alert*

Note: This tool will trigger a suicide warning when a client's mood, thoughts, or behaviors are indicative of possible suicidality. It is possible that the CAT-SS may trigger a suicide warning even if the C-SSRS does not because it seeks to understand suicidality more broadly; the CAT-SS may trigger a warning for individuals with severe suicidal ideation but no plan, intent, or behavior (which would be missed by the C-SSRS).

DEPRESSION

The CAT-MH® includes two modules to measure depression - the CAD-MDD (Major Depressive Disorder) and the CAT-DI (Depression Inventory). The differences between these modules are outlined below.

BRIEF DIAGNOSTIC SCREENER FOR MAJOR DEPRESSIVE DISORDER: CAD-MDD

The CAD-MDD is a brief screener for Major Depressive Disorder (MDD) requiring, on average, 4 questions to predict the results of a Structured Clinical Interview for DSM Disorders (SCID) for DSM-5 MDD. The accuracy of the CAD-MDD can be understood in two ways -- through its high sensitivity¹ and its high specificity². The sensitivity of the CAD-MDD is 95% and the specificity of the CAD-MDD is 87%. This module provides a binary diagnosis decision (positive or negative) and a measure of confidence (between 0% to 100%) for the diagnosis. The CAD-MDD is best used as a screener for MDD; it has limited value for measurement over time unless the interest lies in understanding whether an individual meets the criteria for MDD.

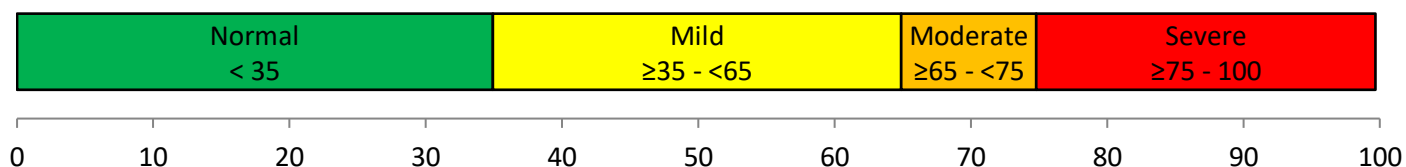
Metrics: *Diagnosis, Confidence*

Note: The confidence measure is between 0% to 100%, where 0.001% is the least confident and 99.999% is the most confident that this individual would receive a diagnosis of MDD if a full diagnostic interview was given.

DEPRESSION INVENTORY: CAT-DI

The CAT-DI effectively measures depression severity based on a client's experience of symptomatology. With an average of ten questions, the CAT-DI maintains a high correlation ($r=0.95$) with the entire bank of 389 depression questions and has been validated against the SCID for the DSM-5. The CAT-DI is best used for screening and measuring depression and is an ideal tool for measurement and evaluation over time.

Severity Score Thresholds



Metrics: *Severity, Category, Precision, Prob of MDD, Percentile, PHQ-9 Equivalent*

Additional Metrics Explained:

¹ *Sensitivity* refers to the CAD-MDD's ability to correctly identify an individual who meets the criteria of MDD as positive. The CAD-MDD is a highly sensitive test which means that there are very few cases where a client who meets the criteria of MDD is not identified by the module.

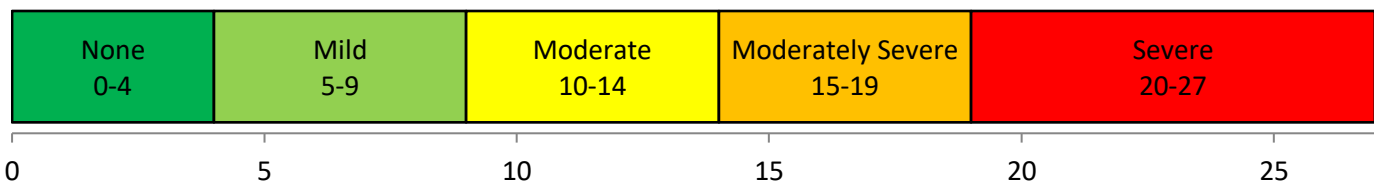
² *Specificity* refers to the CAD-MDD's ability to identify an individual who does not meet the criteria of MDD as negative. The CAD-MDD, along with being a highly sensitive test, is also a highly specific test. There is a relatively low risk of a false positive (someone who does not actually meet the criteria of MDD being identified as someone with MDD).

- **Prob of MDD:** the probability that, given the severity score produced by the CAT-DI, a client would receive a diagnosis of MDD if a full diagnostic interview was given. This number will always be between 0 and 1, where 1 is the highest probability.
- **Percentile:** the percent of the population with a clinician-based diagnosis of MDD whose severity score is less than the client's. This number will always be between 0 and 100. If a client has a percentile score of 99, it means that 99% of all clients who have received the assessment and were diagnosed with MDD based on a structured clinical interview had a lower severity score.
- **PHQ-9 Equivalent³:** The PHQ-9 equivalent score translates the results of the CAT-DI into the equivalent PHQ-9 score for people who are familiar with the PHQ. We do **not** recommend using traditional PHQ-9 thresholds to evaluate the CAT-DI score.

THE PHQ-9

The PHQ-9 (Patient Health Questionnaire) is a 9-item screening measure for the severity of depression. It incorporates DSM-IV depression diagnostic criteria with other leading major depressive symptoms into a brief self-report tool. The PHQ-9 is not a CAT-MH[®] module and is included only for the convenience of organizations that require it for reporting purposes.

Severity Thresholds



Metrics: Severity, Category

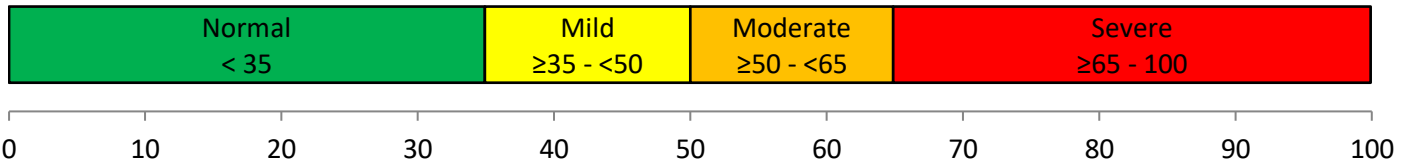
Notes:

1. The PHQ-9 is much less precise than the CAT-DI, especially when assessing clients with lower levels of depression.
2. The CAD-MDD and CAT-DI, like the PHQ-9, are approved for use as a part of the National Quality Forum NQF-0418 standard for annual depression screening.

³ The correspondence between the CAT-DI based PHQ-9 equivalent and the actual PHQ-9 score increases with increasing severity because there is more variability for the PHQ-9 for lower scores. This is not true for the CAT-MH[®] measures, like the CAT-DI, where items are administered until a constant level of precision is achieved. Note also that the CAT-DI is correlated $r=0.81$ with the PHQ-9 (see Gibbons et.al. 2012), and the test-retest reliability for the PHQ-9 is $r=0.84$ (Kroenke et.al. 2001). As such the difference between the CAT-DI estimated PHQ-9 equivalent score and the actual PHQ-9 score is similar to the difference between two actual PHQ-9 scores for the same person. Note that the test-retest reliability for the CAT-DI is higher than the PHQ-9 ($r=0.92$), despite the fact that the CAT-DI asks different questions upon repeat administration. This is due to the increased precision of measurement of the CAT-DI.

CAT-ANX

The CAT-ANX effectively measures anxiety severity based on a client's experience of symptomatology. With an average of 10 questions, the CAT-ANX maintains a high correlation ($r=0.94$) with the entire bank of 431 anxiety questions and has been validated against the SCID for the DSM-5. Like the CAT-DI, the CAT-ANX is best used for screening and measuring anxiety and is ideal for measurement and evaluation over time.

Severity Score Thresholds

Metrics: *Severity, Category, Precision, Prob of GAD, Percentile*

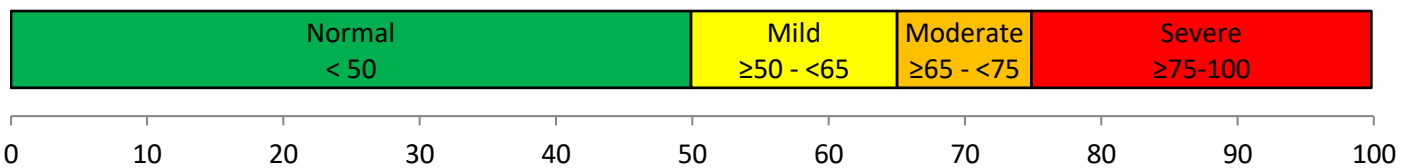
Additional Metrics Explained:

- **Prob of GAD:** the probability that, given the severity score produced by the CAT-DI, a client would receive a diagnosis of Generalized Anxiety Disorder if a full diagnostic interview was given. This number will always be between 0 and 1, where 1 is the highest probability.
- **Percentile:** the percent of the population with a clinician-based diagnosis of GAD whose severity score is less than the client's. This number will always be between 0 and 100. If a client has a percentile score of 99, it means that 99% of all clients with a GAD diagnosis who have received the assessment had a lower severity score.

CAT-MANIA

The CAT-MANIA effectively measures mania/hypomania severity based on a client's experience of symptomatology. With an average of 12 questions, the CAT-MANIA maintains a high correlation ($r=0.91$) with the entire bank of 87 questions related to mania/hypomania, and the CAT-MANIA has been validated against SCID DSM-5 diagnoses for current bipolar I and II. The CAT-MANIA is best used for screening and measuring active mania/hypomania and is an ideal tool for measurement and evaluation over time. Although validated against clinician-based bipolar diagnoses, the CAT-MANIA should not provide a Bipolar Disorder diagnosis. Rather, high scores should prompt the clinician to consider further evaluation to determine if the patient has bipolar disorder. The CAT-MANIA can then be used to monitor changes in the severity of the bipolar disorder.

Severity Thresholds



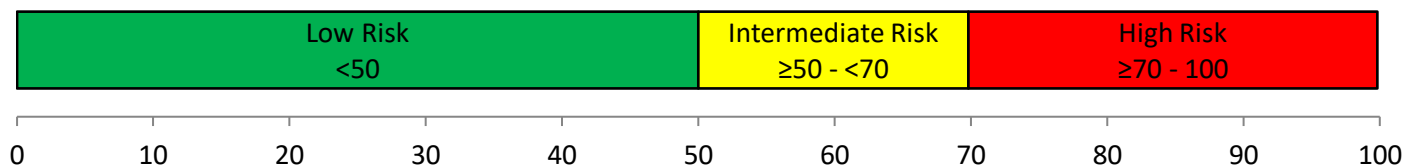
Metrics: Severity, Category, Precision

Note: The CAT-MANIA should not be used to predict Bipolar Disorder because it is a lifetime diagnosis. A patient may be neither depressed nor manic during the past 2 weeks and still meet the criteria for Bipolar Disorder. Elevated scores should suggest further study of the possibility of bipolar disorder. The CAT-MANIA can be used to measure changes in the severity of bipolar disorder in patients with a clinically confirmed bipolar disorder.

SUBSTANCE USE DISORDER

The CAT-MH® includes two modules related to Substance Use Disorder (SUD) risk - the CAT-SUD and the CAT-SUD-E - which collect different types of SUD data (detailed below). Despite their differences, they share thresholds and constraints.

Severity Thresholds



Notes:

1. The CAT-SUD and CAT-SUD-E are only available for use in assessing SUD risk over the past 30 days or the past year; no other time frames are available. The CAT-SUD-E also provides diagnostic screening for lifetime SUDs.
2. The CAT-SUD and CAT-SUD-E adaptively measure the *risk* of SUD based, in part, on social support and mood factors. Because of this, it is possible for a person who is not using substances to have an elevated severity score -- especially if they have high levels of comorbid depression, anxiety, and/or PTSD.
3. The CAT-SUD-E includes the adaptive severity scale from the CAT-SUD, but adds a diagnostic screening component for 7 SUDs for adults and 9 for youth.

CAT-SUD FOR LONGITUDINAL USE

The CAT-SUD is the first adaptive measure for substance use disorder. It provides a crosswalk between mental health symptoms (depression, anxiety, PTSD), social supports, risky behaviors, and self-reports of the abuse of substances (tobacco, alcohol, illegal drugs, opioids). It can detect Substance Use Disorder (SUD) risk in people who have not yet started to abuse substances or refuse to admit that they are. The CAT-SUD provides a severity score that refers to the risk of substance use disorder as well as self-reports of the use/abuse of specific substances. The test has been validated against the CIDI. The CAT-SUD is best used for screening and measuring substance use disorder risk and actual substance use and is an ideal tool for measurement and evaluation over time.

Metrics: *Severity, Category, Precision, Alcohol, and frequency of use of Sedatives/Hypnotics, Opioids/Analgesics, Heroin/Methadone, and Cocaine/Amphetamines during the past 30 days, or a 30-day period during the past year.*

Additional metrics: All substance-related metrics (Alcohol, Sedatives/Hypnotics, etc.) identify how many days in the last month a client reported using a specific substance.

Note: The CAT-SUD does not provide a diagnosis of SUD – it provides a SUD severity score. Note that even in subjects that do not have a SUD, their severity score may be non-zero because they have precursors that put them at increased risk of developing a SUD in the future.

CAT-SUD-E FOR DIAGNOSTIC SCREENING

The CAT-SUD-E collects similar, though not identical data, to the CAT-SUD. Unlike the CAT-SUD, the CAT-SUD-E does not include self-reports related to frequency of use in the past 30 days for specific substances, and instead, provides overall and substance-specific diagnoses. The result is that the CAT-SUD-E is able to provide diagnoses (current or lifetime) for specific substance use disorders related to alcohol, marijuana, nicotine, opioids, cocaine, amphetamines, and hallucinogens. The CAT-SUD-E is best used for screening and measuring specific substance use disorders and for measuring the risk of substance use at a given time. The CAT-SUD is more applicable for repeat assessments. Many institutions use the CAT-SUD-E initially for diagnostic screening and severity measurement, and then follow-up using the CAT-SUD for longitudinal severity assessments and frequency of use assessments.

Metrics: *Severity, Category, Precision, Diagnoses (substance and timeframe [current versus lifetime])*

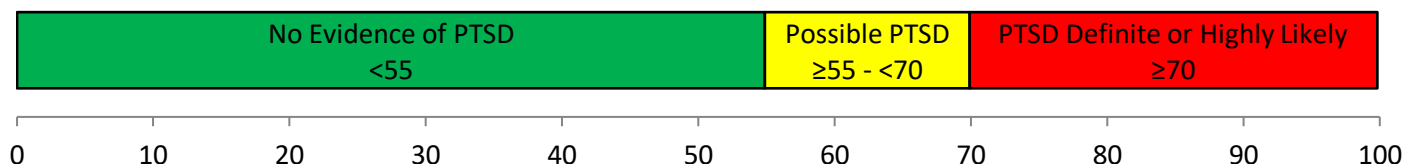
Additional metrics: All diagnosis-related metrics will be reported with both the specific substance disorder and the information as to whether it refers to a current disorder, or lifetime disorder.

Note: The CAT-SUD-E does not capture frequency of use in a given month. This information is only available within the CAT-SUD. The CAT-SUD-E has been validated against the SCID for DSM-5 substance use disorders.

CAT-PTSD FOR LONGTUDINAL USE [RESTRICTED, REPLACED WITH PTSD-E ON 12/19/2021]

The original CAT-PTSD effectively measures PTSD severity based on a client's experience of symptomatology. With an average of 6 questions, the CAT-PTSD maintains a high correlation ($r=0.98$) with the 20-item PCL-5. The CAT-PTSD is best used for screening and measuring current PTSD severity and is an ideal tool for measurement and evaluation over time.

Severity Thresholds



Metrics: Severity, Category, Precision

PTSD-DX FOR SCREENING

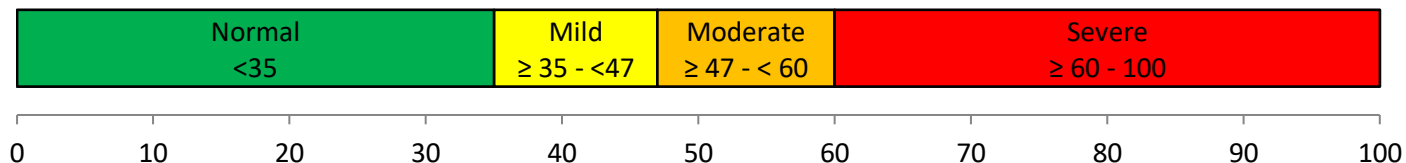
The CAD-PTSD (“PTSD-DX”) is used for diagnostic screening. It reproduces a CAPS-5 structured clinical interview with a trained clinician with $AUC=0.91$ (outstanding range) using a maximum of 6 items in a median time of 35 seconds. To further reduce burden, conditional testing is available with the PTSD-DX and PTSD-E. When ‘If Positive PTSD-Diagnosis’ is selected the PTSD-DX module will be administered first and if the respondent screens positive for PTSD, the PTSD-E will be administered to establish a severity score with clinically meaningful thresholds of mild, moderate, and severe. This reduces median administration time to 35 seconds for most people, and 94 seconds for those who screen positive.

Metrics: Diagnosis, Probability of PTSD

PTSD-E FOR SEVERITY MEASUREMENT AND LONGITUDINAL ASSESSMENTS

The CAT-PTSD-EXPANDED (“PTSD-E”) is used for measurement-based care, treatment triage, and outcome assessment. It produces a continuous severity score on a scale of 0 to 100 with approximately 5 points of precision. Total administration time is an average of 94 seconds. These tests were calibrated and validated in the Veterans Administration, and the item bank was developed to apply to both military and civilian forms of PTSD. Convergent validity of the PTSD-E was demonstrated against the PCL-5 ($r=0.88$) and diagnostic validity against the CAPS-5 ($AUC=0.85$ – excellent range).

Severity Thresholds

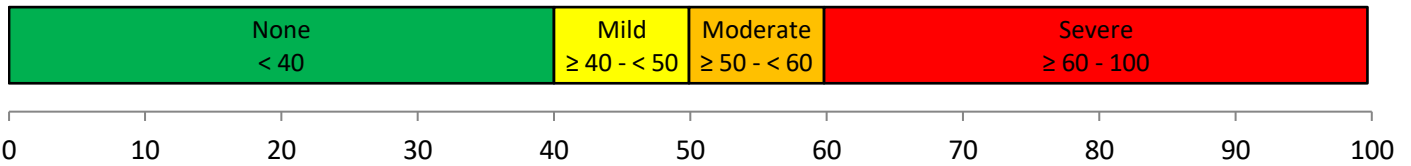


Metrics: Severity, Category, Precision

CAT-PSYCHOSIS

The CAT-PSYCHOSIS measures the severity of psychosis and can be administered either as a self-report or a clinician-reported assessment. Both versions of the CAT-PSYCHOSIS have been validated against the SCID for DSM-5 (for current psychosis). Both versions require an average of 11 questions and maintain a high correlation ($r=0.94$) with the entire bank of 73 questions related to psychosis. On average, a self-report version of the CAT-PSYCHOSIS requires an average of 90 seconds to complete, and the clinician-reported version requires an average of 5 minutes. The CAT-PSYCHOSIS is best used to screen for and measure active psychosis and to measure the severity of psychosis over time.

Severity Thresholds



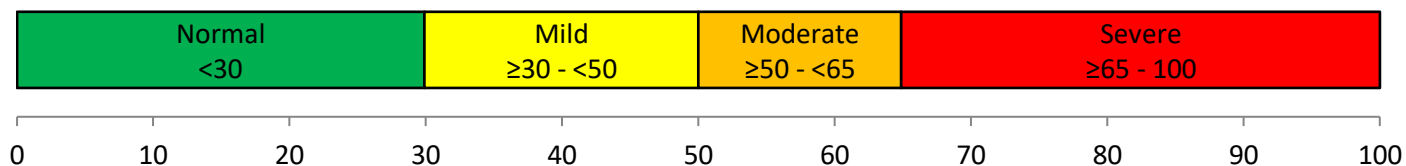
Metrics: *Severity, Category, Precision*

Note: The CAT-PSYCHOSIS should not be used to diagnose psychotic disorders (such as schizophrenia) because these are considered a lifetime diagnosis. Because of this, a person may not be actively experiencing psychosis while meeting the criteria for a psychotic disorder, and conversely, someone may experience a period of psychosis without meeting the more robust criteria.

CAT-ADHD

The CAT-ADHD measures the severity of attention-deficit/hyperactivity disorder symptomatology for adults 18 and over and takes an average of less than 2 minutes to complete. The CAT-ADHD is an adult version of the K-CAT® module for ADHD, which was validated against the K-SADS clinician diagnosis. The CAT-ADHD requires an average of 7 questions and maintains a correlation of $r=0.95$ with the entire bank of 153 questions related to ADHD. The CAT-ADHD is best used for screening and measuring current ADHD symptom severity and is an ideal tool for measurement and evaluation over time.

Severity Thresholds

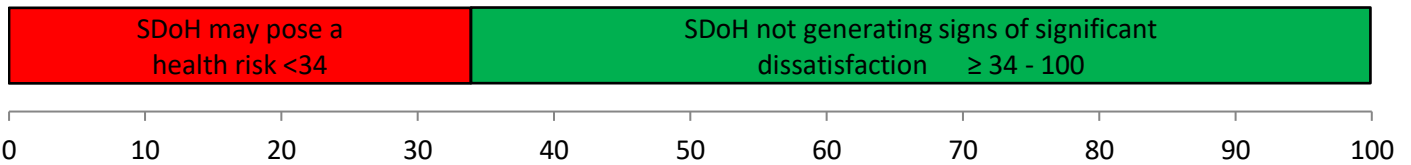


Metrics: Severity, Category, Precision

CAT-SDOH

The CAT-SDoH measures a person's subjective perspective of their social and physical environment in relation to their health. This module requires, on average, 9 questions to maintain a high correlation ($r=0.95$) with the entire bank of questions related to Social Determinants of Health. The CAT-SDoH module is best used to screen and measure an individual's general satisfaction with their social and physical environment and can be used to measure change over time; changes in the score over time represent changes in a person's perception about the quality of their life.

Severity Thresholds



Metrics: *Severity, Category, Precision*

Notes:

1. Higher scores in CAT-SDoH represent higher levels of satisfaction, and so a higher severity score is a positive indicator.
2. This module consists of seven subdomains (Family, Finance, Health, Leisure, Living, Safety, and Social) and uses a 7-point scale for each question with options ranging from terrible (1) to delighted (7).
3. The CAT-SDoH can also be used as a quality of life measure.