ADAPTIVE TESTING TECHNOLOGIES | UNRIVALED MEASUREMENT
VALIDATED • PRECISE • ACCURATE

CAT-MH® INTERPRETIVE GUIDE

GUIDELINES FOR A GENERAL POPULATION

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.

CONTENTS

THE VALUE OF THE CAT-MH®

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, among adults ages 18 and over, including:

- Depression
- Anxiety
- Suicidality
- Mania/Hypomania
- Substance Use Disorder
- Post-Traumatic Stress Disorder
- Psychosis
- Adult ADHD
- Social Determinants of Health

The full battery of modules can be administered to a client in 10 to 15 minutes on any web-enabled device via Pearson's Q-global® platform. However, many users target specific modules to their population, resulting in a comprehensive assessment completed in even shorter 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® METRICS

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. This allows the clinician to assess the clinical and statistical significance of change at the level of the individual at any point in time.

The value of these metrics lies in their ability to help a clinician understand the type and degree of psychopathology a client is experiencing at a particular point in time and the degree of uncertainty in that knowledge. This precision has important implications in practice because it allows you to measure change over time in a statistically meaningful way, at the individual client level.

The severity score and accompanying category tell the clinician how a given client is presenting *in relation to* the general population tested with that module. 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 within the interval of 65 to 85, 95% of the time (i.e., 95% confidence interval = severity score ± 1.965*precision). Second, precision can be used to determine if an individual's change over time is clinically and / or statistically significant.

CAT-MH® categorizes change in three ways: uncertain, clinically significant, and statistically significant. Table 1 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 *at least* 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.

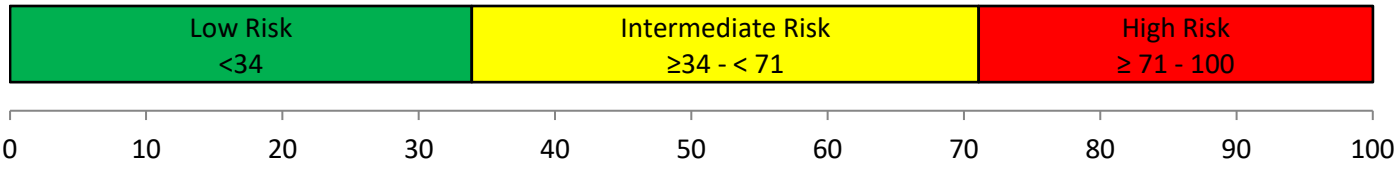
With this data, clinicians can assess an individual's status on a monthly, weekly, or daily basis and adjust care as needed. Longitudinal analysis of CAT-MH® scores used in treatment planning enables evaluation of the effectiveness of different modalities and strategies. Conventional mental health measurement scales lack uncertainty estimates and therefore cannot accurately measure change at the individual client level.

SUICIDE: CAT-SS

The CAT-SS is a dimensional severity measure designed to assess suicidal thoughts, intent, plans, and behaviors. It draws items from depression, anxiety, and suicidal ideation domains that collectively define a single unified dimension, providing a crosswalk between depression and anxiety precursors to suicidality. On average, it requires approximately 10 items to effectively assess for suicidality and maintains a strong correlation (r = 0.96) with the full 111-item bank. The CAT-SS is best used to measure current 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.

The CAT-SS has been validated against the clinician-administered Columbia-Suicide Severity Rating Scale (C-SSRS), demonstrating sensitivity of 1.00 and specificity of 0.95 for identifying suicidal ideation. Prospective studies indicate that CAT-SS scores predict future suicide attempts and suicidal ideation with intent and/or plan, and that the measure is sensitive to change in both small and large randomized clinical trials.

Severity Score Thresholds



Metrics: *Severity, Category, Precision, Alert*

Note: This tool will trigger a suicide warning on the Q-global® Score Report when a client's mood, thoughts, or behaviors are indicative of possible suicidality. When the Suicide Module is administered, it will appear first on the report and should be reviewed immediately after completion of the testing.

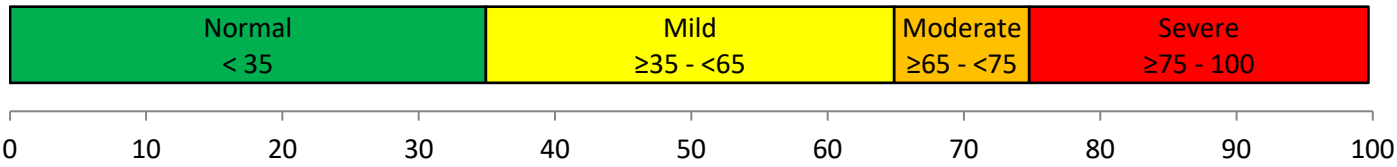
DEPRESSION: CAT-DI

The CAT-DI effectively measures depression severity based on a client's experience of symptomatology. It draws from a 389-item bank and administers an average of 10-12 questions while maintaining a high correlation (r = 0.95) with the full item bank (r = 0.95). CAT-DI has been validated against the Structured Clinical Interview for DSM-5 (SCID-5) and is approved for use under the National Quality Forum (NQF) 0418 measure for annual depression screening.

The CAT-DI provides a continuous depression severity score, an estimated probability of meeting DSM-5 criteria for Major Depressive Disorder (MDD), and a percentile rank relative to individuals who meet criteria for MDD. To support continuity with existing workflows, CAT-DI provides a PHQ-9 equivalent score alongside its more precise severity estimate. CAT-DI correlates strongly with the PHQ-9 (r = 0.81) and demonstrates higher test–retest reliability (r = 0.92 versus r = 0.84 for the PHQ-9), despite not repeating the same items across administrations. Because the PHQ-9's test–retest correlation (r = 0.84) is similar to its correlation with CAT-DI (r = 0.81), differences between a CAT-DI–derived PHQ-9 equivalent score and an observed PHQ-9 score are comparable to the typical variation between two PHQ-9 administrations for the same patient.

The CAT-DI is best used to screen for depression, measure severity, and to monitor and evaluate changes in depression over time.

Severity Score Thresholds



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

Additional Metrics Explained:

- **Probability 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 Equivalent**[1]: To support continuity with existing workflows 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 the equivalent score is included only for the convenience of organizations that require it for reporting purposes.

Notes:

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

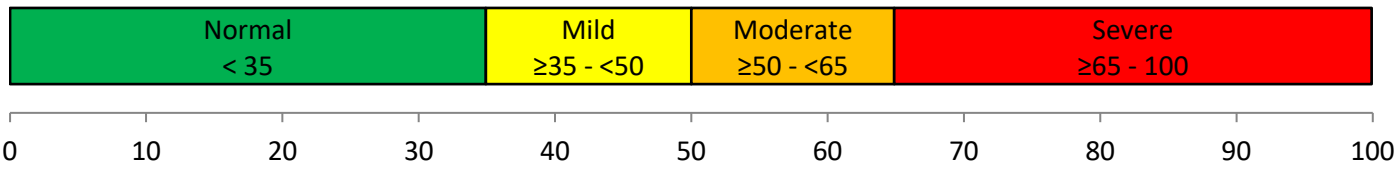
[1] The correspondence between the CAT-DI based PHQ-9 equivalent and the actual PHQ-9 score increases with increasing severity because there is more variability for the PHQ-9 for lower scores. This is not true for the CAT-MH® measures, like the CAT-DI, where items are administered until a constant level 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.

ANXIETY DISORDER: CAT-ANX

The CAT-ANX effectively measures anxiety severity based on a client's experience of symptomatology and provides a continuous anxiety severity score. It maintains a high correlation with the full 431-item bank (r = 0.94) while administering, on average, 10-12 items and typically takes about 79 seconds to complete. Validation studies used the SCID-5 to obtain diagnostic classifications for Generalized Anxiety Disorder (GAD) and Major Depressive Disorder, and examined convergent validity with established measures, showing correlations of r = 0.73 with the HAM-D, r = 0.78 with the PHQ-9, and r = 0.81 with the CES-D. Diagnostic performance for identifying GAD demonstrated a sensitivity of 0.65 and specificity of 0.93; the lower sensitivity reflects the DSM-5 requirement that GAD symptoms persist for at least six months, meaning patients with high anxiety that has not yet lasted six months may score high on CAT-ANX but not receive a GAD diagnosis. Participants were recruited from a psychiatric clinic and a community mental health center, supporting applicability across clinical settings.

The CAT-ANX is best used for screening and measuring anxiety and is ideal for measurement and evaluation over time. Use the continuous severity score to screen and quantify current anxiety and to monitor change over time; interpret the estimated probability of meeting DSM-5 criteria for GAD as a risk indicator rather than a standalone diagnosis; and use the percentile rank to contextualize severity relative to individuals who meet criteria for GAD.

Severity Score Thresholds



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

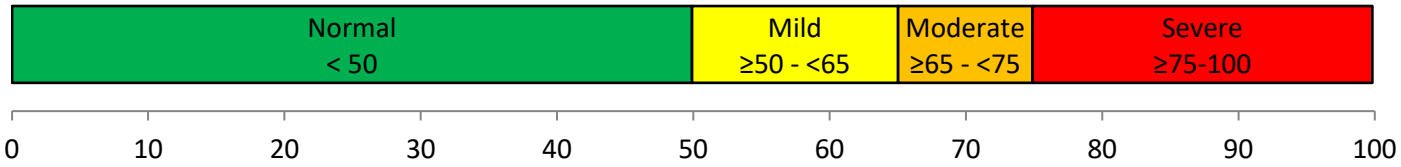
Additional Metrics Explained:

- **Probability 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.

MANIA/HYPOMANIA: 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-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 does 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 Score Thresholds



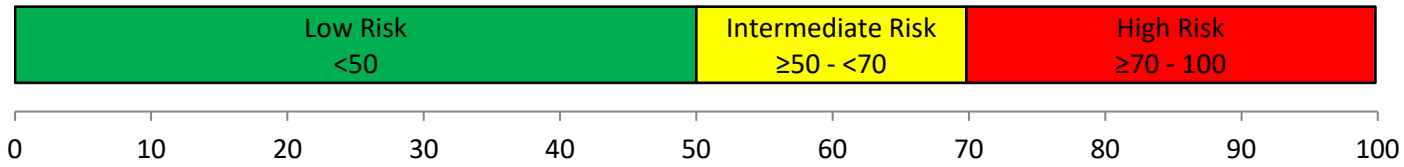
Metrics: *Severity, Category, Precision*

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® has two modules for assessing Substance Use Disorder (SUD) risk: CAT-SUD and CAT-SUD-E. Each gathers different data (detailed in the following sections) but uses the same thresholds and constraints.

Severity Score 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 8 SUDs.

SUBSTANCE USE DISORDER – INITIAL ASSESSMENT: CAT-SUD-E

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 can provide diagnoses (current or lifetime) for specific substance use disorders related to alcohol, marijuana, nicotine, opioids, cocaine, amphetamines, hallucinogens, and sedatives. 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.

This report has two parts. The first provides the client's self-reported current and past substance use, with a comprehensive set of lifetime and current SUD diagnoses. The second part provides the typical CAT-MH® metrics of severity, category, and precision regarding the individual's current substance use disorder severity.

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-5 substance use disorders.

SUBSTANCE USE DISORDER – OUTCOME MONITORING: CAT-SUD

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) in the past 30 days. It can detect 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 Composite International Diagnostic Interview (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 individuals 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.

PTSD DIAGNOSIS: PTSD-DX

The CAD-PTSD (PTSD-DX) is used for diagnostic screening that quantifies PTSD severity on a 0-100 scale. It reproduces a Clinician-Administered PTSD Scale for DSM-5 (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. Unlike the other computer adaptive tests (CAT) in the CAT-MH® suite, the CAD-PTSD is a computer adaptive diagnostic test (CAD). This means it uses decision trees and random forests to minimize the number of items needed to shift from a negative diagnosis to a positive one. Its purpose is to screen for an underlying disorder (PTSD) as opposed to determining a dimensional severity measure. Therefore, the metrics provided are a diagnosis (positive or negative) and a probability of PTSD.

Metrics: *Diagnosis, Probability of PTSD*

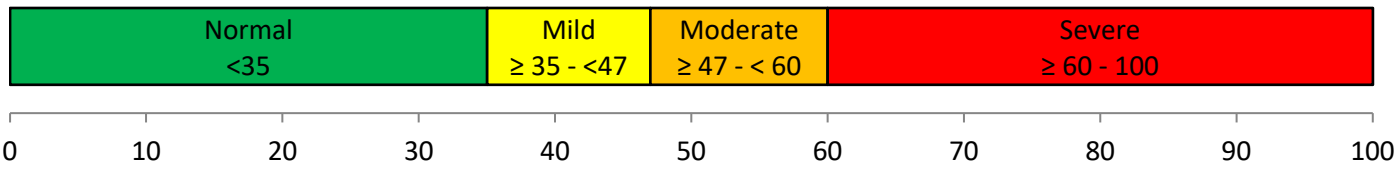
PTSD EXPANDED: PTSD-E

The CAT-PTSD-EXPANDED (PTSD-E) is a dimensional measure that reports PTSD severity on a continuous 0-100 scale with approximately 5 points of precision. It is optimal for measurement-based care, treatment triage, and outcome assessment. Using adaptive testing, it reproduces the full 203-item bank score with high fidelity (r = 0.95) while administering an average of 10 items in approximately 94 seconds, reducing client burden by 95%. Convergent validity with the PCL-5 is strong (r = 0.88), and diagnostic validity against the CAPS-5 is in the excellent range (AUC = 0.85).

The continuous severity scores on the CAT-PTSD-E map to four severity categories: none, mild, moderate, and severe. The shift between none and mild was selected to have high sensitivity and moderate specificity, mild vs. moderate to have high sensitivity and high specificity, and moderate vs. severe to have high specificity. These decision rules yielded thresholds of 35 (none vs. mild: 95% sensitivity, 51% specificity), 47 (mild vs. moderate: 79% sensitivity, 78% specificity), and 60 (moderate vs. severe: 50% sensitivity, 93% specificity). While the continuous score should be used for measurement and tracking change over time, these categories can aid clinical decision-making. Validation was conducted among U.S. military veterans within the Veterans Health Administration, and the item bank was developed for applicability to both military and civilian PTSD.

Note that the CAT-PTSD and CAT-PTSD-E modules cannot be administered within the same test session.

Severity Score Thresholds

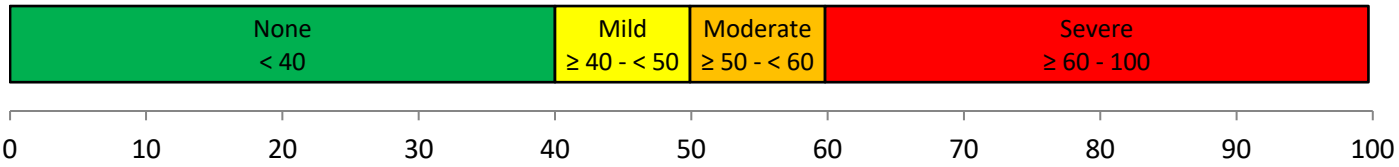


Metrics: *Severity, Category, Precision*

PSYCHOSIS – SELF-REPORT: CAT-PSYCHOSIS

The CAT-PSYCHOSIS provides a dimensional assessment of current psychosis severity based on client self-report. It uses adaptive administration to select an average of 11 items while maintaining strong fidelity to the full 73-item bank (r = 0.94) and can typically be completed in under two minutes (about 90 seconds on average). Scores are reported on a 0–100 scale with approximately 6-point precision, supporting sensitive measurement of symptom change. Validity has been demonstrated against the SCID-5 Psychotic Disorder diagnosis and the clinician-rated Brief Psychiatric Rating Scale (BPRS) in samples that include inpatient and outpatient psychiatric patients as well as healthy controls. The module effectively discriminates psychosis from healthy status (AUC=0.850). The CAT-PSYCHOSIS is best used to screen for and measure active psychosis and to measure the severity of psychosis over time.

Severity Score Thresholds



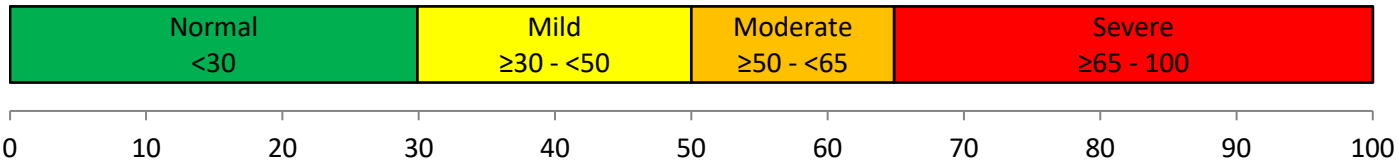
Metrics: *Severity, Category, Precision*

Note: The CAT-PSYCHOSIS should not be used to diagnose psychotic disorders (such as schizophrenia) because these are considered lifetime diagnoses. 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.

ADULT ADHD: CAT-ADHD

The CAT-ADHD measures attention-deficit/hyperactivity disorder symptom severity for adults 18 and over. It is an adult version of the K-CAT® module for ADHD, which was validated against the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS) clinician diagnosis. The CAT-ADHD requires an average of seven questions, maintaining a strong correlation with the full 153-item ADHD bank (r = 0.95), with an average administration time of under two minutes (approximately 109 seconds). This efficiency minimizes client burden without sacrificing measurement fidelity. The CAT-ADHD is best used for screening and measuring current ADHD symptom severity and is an ideal tool for supporting measurement-based care and longitudinal evaluation of ADHD symptomatology.

Severity Score Thresholds

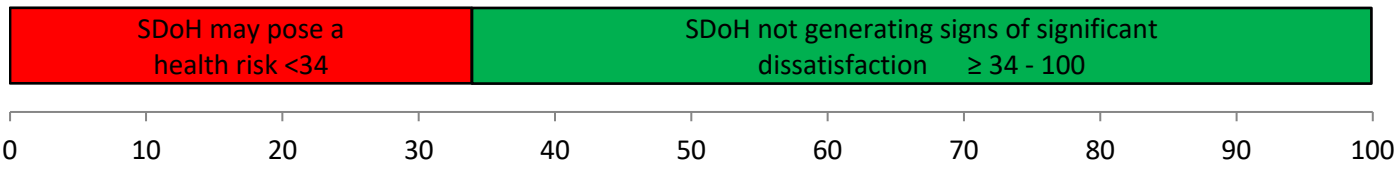


Metrics: *Severity, Category, Precision*

SOCIAL DETERMINANTS OF HEALTH: CAT-SDOH

The CAT-SDoH is an adaptive version of a life quality measure specifically designed and validated for people with mental health disorders. It offers a comprehensive, dimensional assessment of an individual's subjective view of their social and physical environment as it pertains to their overall health. The original scale comprises seven subdomains—Family, Finance, Health, Leisure, Living, Safety, and Social, with 4-6 items per domain (35 items total). Each item is rated on a 7-point scale from Terrible (1) to Delighted (7), with a Not Applicable option available for some questions. Using adaptive administration, CAT-SDoH requires an average of 11 items while maintaining a strong correlation with the full 35-item score (r = 0.96). Results are reported on a 0–100 scale with approximately 5-point precision. Higher scores indicate overall satisfaction, and lower scores indicate dissatisfaction, with scores below 34 flagging elevated dissatisfaction. Specific item responses help identify which aspects of the environment contribute to the overall score. The CAT-SDoH module is best used to screen and measure general satisfaction with their social and physical environment and to monitor change over time, where changes in the score reflect shifts in perceived life quality. It can be compared or integrated with traditional SDoH measures to provide a more complete picture, and substantial shifts in SDoH severity scores may signal underlying mental health concerns.

Severity Score Thresholds



Metrics: *Severity, Category, Precision*