



HHS Public Access

Author manuscript

J Am Acad Child Adolesc Psychiatry. Author manuscript; available in PMC 2021 November 01.

Published in final edited form as:

J Am Acad Child Adolesc Psychiatry. 2020 November ; 59(11): 1264–1273. doi:10.1016/j.jaac.2019.08.009.

Computerized Adaptive Tests for Rapid and Accurate Assessment of Psychopathology Dimensions in Youth

Robert D. Gibbons, PhD,

University of Chicago, IL.

David J. Kupfer, MD,

University of Pittsburgh School of Medicine, PA.

Ellen Frank, PhD,

University of Pittsburgh School of Medicine, PA.

Benjamin B. Lahey, PhD,

University of Chicago, IL.

Brandie A. George-Milford, MA,

UPMC Western Psychiatric Hospital, Pittsburgh, PA.

Candice L. Biernesser, LCSW, MPH,

UPMC Western Psychiatric Hospital and the University of Pittsburgh Graduate School of Public Health, Pittsburgh, PA.

Giovanna Porta, MS,

UPMC Western Psychiatric Hospital, Pittsburgh, PA.

Tara L. Moore, MA, MPH,

HealthRhythms, Inc., New York, NY.

Jong Bae Kim, PhD,

Center for Health Statistics, University of Chicago, IL.

David A. Brent, MD

University of Pittsburgh School of Medicine, PA.

UPMC Western Psychiatric Hospital, Pittsburgh, PA.

Abstract

Background—At least half of youth with mental disorders are unrecognized and untreated. Rapid, accurate assessment of child mental disorders could facilitate identification and referral and potentially reduce the occurrence of functional disability that stems from early-onset mental disorders.

Correspondence to Robert D. Gibbons, PhD, University of Chicago, 5841 S. Maryland Ave., Room W260, MC2000, Chicago, IL 60637; rdg@uchicago.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Method—Computerized adaptive tests (CATs) based on multidimensional item response theory were developed for depression, anxiety, mania/hypomania, attention deficit hyperactivity disorder, conduct disorder, oppositional defiant disorder, and suicidality, based on parent and child ratings of 1060 items each. In Phase 1, CATs were developed from 801 participants. In Phase 2, predictive, discriminant, and convergent validity were tested against semi-structured research interviews for diagnoses and suicidality in 497 patients and 104 healthy controls. Overall strength of association was determined by area under the receiver operator curve (AUC).

Results—The child and parent independently completed the Kiddie-CATs (K-CATs) in a median time of 7.56 and 5.03 minutes, respectively, with an average of 7 items per domain. The K-CATs accurately captured the presence of diagnoses (AUC's from 0.83 for Generalized Anxiety Disorder to 0.92 for Major Depressive Disorder) and suicidal ideation (AUC=0.9966). Strong correlations with extant measures were found ($r = 0.60$). Test-retest reliability averaged $r=0.80$.

Conclusion—These K-CATs provide a new approach to child psychopathology screening and measurement. Testing can be completed by child and parent in less than 8 minutes and yields results that are highly convergent with much more time-consuming structured clinical interviews and dimensional severity assessment and measurement. Testing of the implementation of the K-CAT is now indicated.

Keywords

adaptive testing; multidimensional item response theory; child and adolescent psychopathology; measurement; diagnosis

Introduction

Mental disorders in children and adolescents are widespread,¹ with over 1/5 of US adolescents having had a psychiatric disorder associated with severe impairment at some point in their lifetime.² However, at least half of youth with mental disorders are untreated.^{3,4} Mental disorder in childhood sets in motion a developmental cascade that leads to functional impairment, which in turn leads to additional psychological symptoms and disease progression that can eventuate in long-term impairment in adulthood.^{5–8} Since nearly 3/4 of adults with mental disorders report an onset of illness prior to age 18,^{9,10} early identification and treatment of childhood-onset mental disorders could have long-term public health implications across the lifespan.

The low rate of detection of mental health disorders in youth stems from a shortage of individuals trained to assess youth for mental health disorders, compounded by the fact that youth with mental disorders most frequently present in settings other than specialty mental health, such as primary care, schools and emergency departments.⁴ To help span this unmet need in assessment and treatment, instruments are needed to assist with initial detection of disorder and assessment of severity to assist in clinical decision making and appropriate triage., thus facilitating diagnostically specific evidence-based care^{11–12}. Existing questionnaires have significant limitations for screening (e.g., length) and for monitoring of treatment response, since using the same items in the same order can lead to response bias upon repeat administration¹³. In addition, in extant questionnaires, all items are assumed to

be equally discriminating and contribute equally to the assessment of severity. This assumption is often incorrect, and thus, extant fixed questionnaires trade off standardization for precision. The use of adaptive screening, based on item-response theory, can overcome these limitations, by tailoring the severity of the items to the severity level of the person, directly estimating the precision of measurement, insuring that all patients are measured to the same level of precision, and retaining the ability to assess severity through adaptive administration of a small subset of items targeted to the patient's level of severity.

The study reported herein aimed to develop and validate a suite of seven Kiddie – Computerized Adaptive Tests (K-CATs) for the dimensional and diagnostic measurement of mental health disorders in youth based on the combination of child self-report and parent ratings of the child. The K-CAT tailors the selection of questions to the particular informant and to the age of the child (e.g., certain questions, related to sexuality, are not administered to children under 12 years of age). The time (in milliseconds) that it takes the child to respond is used as a safeguard against children that may simply be answering without listening to or reading the question. We have previously demonstrated the utility of computerized adaptive testing based on multidimensional item response theory (MIRT) for the assessment of mental health disorders in adults.¹³⁻¹⁶ The goal of this study was to determine if we could develop efficient, reliable adaptive measures of the presence and severity of child and adolescent depression, mania, anxiety, ADHD, CD, and ODD. We hypothesized that brief K-CATs could be developed that would accurately identify those with one of these diagnoses, that would show good test-retest reliability, and would correlate with other scalar measures of severity.

Method

Overview

In Phase 1, we developed K-CATs based on reports from item banks of 1060 each for the child and parent). MIRT models were fitted to each of the 6 domains separately for parent and child ratings and for child-rated suicidality. Following the item calibrations, the simulated CAT (based on the complete set of item responses for a given domain, e.g., depression) was used to develop the adaptive tests based on over 1000 different possible combinations of CAT tuning parameters. For each diagnostic domain, we selected the simulated CAT that minimized the number of items and maximized the correlation between the CAT score and the total domain-specific total item bank score. The K-CATs developed herein adaptively select a small, statistically optimal subset of items for each child specifically for each diagnostic domain. In Phase 2, we tested the newly developed K-CATs in a new sample of patients and healthy controls who also underwent diagnostic interviews, in order to test predictive, discriminant, and convergent validity, and test-retest reliability.

Sample

Participants were youth aged 7–17 years old. Both the participants and their parent or legal guardian were English speakers. Children were excluded if they had autism spectrum, intellectual developmental, or a psychotic disorder that would limit their ability to provide accurate self-reports. In Phase 1, item-level data were collected to calibrate the MIRT

models and simulated CATs. We screened 1021 patients, assessed 840, excluded an additional 39 in whom we detected exclusionary criteria, resulting in a test sample of 801 youth, of whom 601 had a mental health disorder. A medical record review determined which mental disorder was primary (meaning the most impairing), specifically; major depressive disorder (MDD, n=137), ADHD (n=115), bipolar disorder with manic symptoms (n=68), anxiety (n=124), ODD (n=103) and CD (n=54). Two hundred (n=200) controls without evidence of psychiatric disorder were added to the calibration sample to ensure that the K-CATs were informative throughout the entire range of severity for each of the adaptive dimensional measures.

The primary diagnosis determined which K-CAT scales were administered to the patient and parent. In 81 cases, project staff had questions about the primary disorder and the records were re-reviewed by a psychiatrist. In 22 participants, review of the medical record by a child and adolescent psychiatrist resulted in a reassignment of the primary condition. While use of clinical diagnoses has obvious limitations, nearly ¼ of participants (24.2%) came from clinics that used Kiddie Schedule for Affective Disorders and Schizophrenia, Present and Lifetime Version (K-SADS-PL)¹⁷ to determine clinical diagnoses, and in Phase 2, when the K-SADS-PL was used to establish clinical diagnoses, correspondence between the clinical diagnosis and that derived from the K-SADS-PL was high, ranging from 72.3% for depression to 91.9% for bipolar disorder.

In Phase 2, we validated the developed K-CATs against research diagnostic interviews (the K-SADS-PL), the Columbia Suicide Severity Rating Scale (C-SSRS)¹⁸ and standard scalar measures for depression (the Mood and Feelings Questionnaire (MFQ)¹⁹ and for anxiety (Screen for Anxiety Related and Emotional Disorders (SCARED)).²⁰ In addition, we assessed the severity of mania (past month and lifetime) using the interviewer-rated Brief Mania Rating Scale (CMRS)²¹ in all participants. Functional impairment was assessed by the interviewer-rated Children's Global Assessment Scale (CGAS).²² We screened 1035 participants, found 637 eligible, and of those, recruited 601. Of 601 participants, 497 were treatment-seeking patients with primary diagnoses of MDD (n= 94) anxiety (n=94), bipolar disorder with manic symptoms (n=55), ADHD (n=97), ODD (n=87), and CD (n=70), based on review of medical records, recruited through Western Psychiatric Institute and Clinic (WPIC), mental health community outreach events, and local clinics and providers. One hundred-four (n=104) were healthy controls, recruited locally, who had a brief Pediatric Symptom Checklist (PSC-17)²³ score below 14 and reported they had not had mental health treatment in the previous two years.

Demographics are displayed in Table 1 and diagnostic characteristics of the sample are presented in Table S1, with diagnostic overlap described in a Venn diagram in Figure S1 and Table S2, available online. In Phase 2, a total of 24 MR's were reviewed by the psychiatrist: 1 to change the primary diagnosis, 8 to confirm the primary diagnosis, and 15 to determine if an exclusionary condition was present (in none was there evidence of such conditions). In Phase 2, all youth were used for validation of the K-CAT against the K-SADS-PL because all primary and secondary diagnoses were utilized.

Procedures

The study was approved by the University of Pittsburgh Institutional Review Board, and informed consent/assent was obtained from parents or legal guardians and participants, respectively. Basic demographic information and diagnoses were obtained from the medical record. The K-CAT items were administered using tablet computers. Research staff were available to assist the participants.

For Phase 1, participants and their parent/caregiver answered approximately 250 items across domains and an additional 100–200 items from their primary diagnostic domain. Control participants and their parent/caregiver were randomly assigned to one of the 6 diagnostic domains.

Item bank development

The items were taken from the 137-item Child and Adolescent Psychopathology Scale (CAPS)²⁴ as well as from the 1008 adult CAT-MH item bank¹⁴ (from a review of over 100 mental health measurement instruments), adapted for youth. The CAPS items cover DSM-IV symptoms from ADHD, ODD, CD, MDD, and anxiety disorders. Items were rewritten to describe a consistent two-week time-frame, to be appropriate for children and adolescents, and to be consistent with DSM-5.²⁵ Items were rated on a 5-point Likert scale with categories of 1=not at all, 2=just a little, 3=somewhat, 4=quite a bit, and 5=very much. A small subset of the items were rated true or false (e.g. “Have you used a weapon that could cause serious injury to someone?”). In total, we constructed banks of 1060 items each for the child’s self-rating and for the parent’s rating of the child. Also, we developed the K-CAT-SS (suicide scale) beginning with the Depression bank of 177 items, from which we identified 65 non-suicide items that had an odds ratio in excess of 2.5 per category on each item’s Likert scale versus a binary indicator of the presence of current suicidal ideation or behavior (past two weeks) based on the 10 suicide items in the item bank (child ratings only), yielding a total of 75 items (65 non-suicide and 10 suicide).

Items were tested for readability using the Flesch-Kincaid reading grade level.²⁶ The overall average reading level was grade 4.5 (SD=2.4, median = 4.4); for children under the age of 11, a research assistant offered to read the questions to the participant.

Item calibration

Given the inherent multidimensionality of the constructs, we calibrated each of the 7 constructs using a full-information item-bifactor model.^{27,28} The bifactor model was originally developed for measurement data by Holzinger and Swineford in 1937²⁹ and extended to IRT for discrete item responses by Gibbons and Hedeker in 1992.²⁷ The bifactor model allows each item to load on a primary dimension and one subdomain of interest. The subdomains are selected in advance based on clinical knowledge of the disorder of interest (e.g. somatic, cognitive, mood, suicidality) and has been used to develop CATs for adult mental disorders.^{13–16} The specific subdomains are listed in Table S3, with example items in Table S4, available online. The bifactor model preserves the advantages of a unidimensional model in terms of being able to be expressed as a single-valued index (the score for the

primary dimension) in the presence of conditional dependencies produced by the sampling of items within specific subdomains.

Separate bifactor models were fitted to each of the 6 domains for the parent and child and the suicidality domain for the child. Numbers of items ranged from 75 for suicidality to 244 for anxiety (average of 168 items per domain and rater). We used data from all 801 subjects to estimate the model parameters for each domain and rater which provides an overall 5:1 ratio of the number of subjects to items that has been shown to be sufficient for accurate item parameter estimation.³⁰ Improvement in fit of the bifactor model over a unidimensional alternative was assessed using a likelihood ratio chi-square statistic. All bifactor models were fitted using the POLYBIF program that is freely available at www.healthstats.org.

The estimated scores are reported on a 0–100-point scale metric with increasing scores denoting increased severity. The primary termination criterion for the K-CAT is based on the uncertainty in the final estimated test score. We selected a termination criterion (posterior standard deviation of the Bayes estimated test score) of 5 points of precision on the 100-point scale (standard error of 0.3 on an underlying unit normal scale). The posterior standard deviations for the estimated severity score have no analogue for total scores based on classical test theory. A second termination criteria based on item information is also used. When the amount of information left in the item bank at a given level of severity is insufficient to attain 5 points of precision, the interview is terminated, and the actual level of precision is reported. This typically happens in the extremes of the severity distribution where the added level of precision is not required since, for example, we know the person is either not depressed or among the most severely depressed.

Validation phase

In Phase 2, participants and their parent/caregiver took the live K-CAT. The median test length was approximately 50 items across the 6 diagnostic domains for the parent and child each) and participants were assessed via semi-structured interviews administered by a trained research clinician using the K-SADS-PL-5.²⁷ KSADS evaluators were blind to the K-CAT findings and other scalar scores. All diagnoses were confirmed by a child psychiatrist. Also in Phase 2, we tested a suicidality scale (the K-CAT-SS), which was then validated against the clinician-rated Columbia Suicide Severity Rating Scale (C-SSRS)¹⁸ in 240 participants. To minimize bias, assessment order (interview first vs. tablet computer) was randomly assigned.

Statistical methods

In Phase 2, the relationship between the K-CAT scores based on child and parent domains and clinical assessments using the K-SADS-PL and C-SSRS was assessed using logistic regression; odds ratios (ORs) and 95% confidence intervals were reported for various effect sizes on the underlying K-CAT 0–100 score distributions. We examined changes of 10, 25 and 50 points in magnitude on the K-CAT domains in terms of the increase in likelihood of the corresponding K-SADS-PL-5 disorder. Overall strength of association was determined by area under the receiver operator curve (AUC) and sensitivity at specificity of 80%. Three-fold cross-validation was used to ensure that different data were used to estimate the

classification functions and determine classification accuracy. Sensitivity analyses included (a) AUC analyses stratified by age and sex of the child, and (b) AUC analyses removing the control subjects. This latter analysis is more conservative, since it addresses the ability of the K-CAT measures to predict diagnoses in a purely treatment-seeking population.

The relationship of specific K-CAT test scores and traditional fixed length tests was assessed using correlation coefficients and 95% confidence intervals. The simultaneous relationship between all 12 K-CAT test scores (parent and child scores not including suicidality) and functional status (CGAS) was determined using multiple regression (entering all 12 K-CAT scores as predictors). Finally, test-retest reliability was assessed using correlational analysis in a subsample of 55 subjects measured twice on the same day (an average of 75 and 78 minutes apart for the child and parent, respectively).

Results

Phase 1

Simulated adaptive testing—Simulated adaptive testing from the complete item responses was used to optimize the tuning parameters of the K-CATs and provide evidence of each K-CAT's ability to reproduce the entire item bank. Correlations of $r > 0.92$ with the K-CAT versus entire item bank scores for all 12 (child and parent) K-CATs with the exception of the child-rated CD ($r = 0.86$); however, for parent-rated CD, $r = 0.95$.

Item calibration on the K-CATs—The bifactor models significantly improved fit over unidimensional alternatives in all cases ($p < 0.0001$). Characteristics of the individual K-CATs are presented in Table S5, available online. In general, K-CATs required between 6 and 8 completed items for the parent and child each in order to achieve precision (posterior standard deviation for the score of that person) of 5 points on a 100-point scale (the primary CAT termination criterion). Overall, we started with 1060 items each for the parent and child and retained 964 for the parent and 903 for the child. Of 75 original items, 64 were retained for the assessment of suicidality.

Phase 2

Administration times—In Phase 2, the child and parent completed the entire suite of K-CATs in a median of 7.56 (interquartile range [IQR]=4.73–12.52) and 5.03 minutes respectively (IQR=3.48–7.70). The median number of completed items across the 7 domains for the child and parent were 51 and 38, respectively. By contrast, the KSADS required a median of 59 minutes (IQR=40.7–80.0) for the child to complete and a median of 60.1 minutes for the parent to complete (IQR=43.9–82.33), with a median combined time for the diagnostic interview of 124.85 minutes (IQR=92.83–158.23).

Correlation between parent and child ratings—The correlation between child self-reports and parent ratings in Phase 2 was statistically significant, but generally modest, ranging from a low of $r = 0.27$ (95% CI, 0.19, 0.34) for depression to $r = 0.44$ (95% CI, 0.37, 0.50) for CD, similar to other reports in the literature.³¹

Validation against semi-structured diagnostic interviews—All the scales were significantly associated with their respective diagnosis (Table 2). Using the maximum of the parent and child scores, for every 10-point difference on the 100-point depression scale (i.e. one-tenth of the range of the scale scores) there was a 3-fold increase in the likelihood of a K-SADS-PL DSM-5 MDD diagnosis (OR=3.16, 95% CI=2.46, 4.05). Similar results were seen for the other disorders and their corresponding KCATs, namely with bipolar (BP) I and II disorder (OR=1.48, 95% CI = 1.14, 1.93), generalized anxiety disorder (GAD) (1.93, 95% CI = 1.61, 2.26), ADHD (OR=2.12, 95% CI = 1.84, 2.43), CD (OR=1.93, 95% CI = 1.63, 2.30), and ODD (OR=2.50, 95% CI = 2.12, 2.94) (see Tables S6 and S7, available online, for 25 and 50-point differences).

Figure 1 present ROC curves for MDD and CD that are based on parent, child and combined ratings. For MDD, the child and parent AUCs were similar (0.84 and 0.89 respectively); however, the combination of parent and child reports increased the AUC to 0.92. For CD, the overall accuracy from the combination of parent and child report was AUC=0.89; however, the parents appeared to rate CD more accurately (AUC=0.88) than did the youths (AUC=0.74).

In terms of predictive accuracy, area under the receiver operator curves (AUC of ROC) were all in the excellent range using the combined set of child and parent K-CAT measures to predict each current K-SADS diagnostic category: MDD, 0.92; bipolar disorder (BP) I and II, 0.84; generalized anxiety disorder, 0.83; ADHD, 0.86; CD, 0.89, and ODD, 0.88 (Table 3). Fixing specificity at 0.80, sensitivity is 0.90 for MDD, 0.74 for BP I or II, 0.71 for GAD, 0.75 for ADHD, 0.81 for CD, and 0.80 for ODD.

The K-CAT-SS (child rating only) provided an AUC of 0.95 for current suicidal ideation based on a C-SSRS structured clinical interview for the presence of suicidal ideation. Combining the K-CAT-SS with all of the other child and parent diagnostic K-CATs, the AUC increased to 0.996. (Figure 2). There were too few current suicidal behavior events to perform an analysis; however, AUC for lifetime suicidal behavior (based on K-CATs for current symptoms) was AUC=0.83.

Validation against specific scales of measurement—We compared the K-CAT depression, anxiety, and mania scales to the MFQ, SCARED, and Brief Mania Rating Scale (MRS).³¹ For the child, correlations were $r=0.72$ (0.68, 0.76) for depression, $r=0.60$ (0.55, 0.65) for anxiety, and $r=0.61$ (0.56,0.66) for mania during the past month and $r=0.59$ (0.53, 0.66) for lifetime mania. For the parent, correlations were $r=0.66$ (0.61, 0.70) for depression, $r=0.68$ (0.63, 0.72) for anxiety, $r=0.69$ (0.65, 0.73) for mania during the past month and $r=0.64$ (0.59, 0.69) for lifetime mania.

Functional status—The relationship between the parent and child K-CAT measures and the CGAS functional status measure was significant (multiple $r=0.66$, $p<0.0001$). Individually, significant predictors included child ratings of depression ($p<0.0001$) and parent ratings of ADHD ($p<0.0001$), anxiety ($p=0.01$), CD ($p=0.0002$), and depression ($p<0.0001$).

Test-re-test reliability—Test-retest reliabilities are displayed for each of the K-CAT scales for child, parent and maximum of child and parent ratings in Table S8, available online. High test-retest reliabilities were found for all measures (average of $r=0.80$). Parent ratings had slightly higher test-retest reliabilities than the child ratings.

Sensitivity analyses: differences by age and sex—We examined the relationship between the combined K-CAT measures and individual KSADS-PL-5 diagnoses stratified by age (younger 7–12 vs. older 13–18 years) and gender (Table S9, available online). In general, use of the combined parent and child measures produced similar classification accuracy in terms of AUC across age and sex strata (average AUC: younger 0.89, older 0.88; male participants 0.85, female participants 0.90). For child self-report K-CAT measures, AUC was lower for male participants and younger children; however, parent ratings did not exhibit these differences. For the combined parent and child measures, AUCs were higher in younger children for MDD (0.96 in younger vs. 0.89 in older children) and BP (0.94 in younger vs. 0.86 in older children).

Validation just within the patient group—To eliminate the possibility that AUCs are potentially overestimated in a mixed sample of treatment-seeking patients and healthy controls, we eliminated the controls and re-performed the AUC analyses. The AUC results were still excellent, although the performance was often slightly lower, and never significantly better than when conducting analyses on the entire sample (Table S10, available online).

Discussion

To our knowledge, this is the first application of adaptive testing that includes both the diagnosis and assessment of severity of child psychopathology. We demonstrated that the K-CATs were accurate measures with respect to both the presence and severity of depression, mania, anxiety, ADHD, ODD, CD and suicidality, and that the K-CATs could be administered in less than 8 and 6 minutes for child and parent, respectively. We also demonstrated test-retest reliability and similar performance across age and sex strata. Although both parent and child report showed convergent validity with respect to diagnosis and severity, the best results were obtained by combining both parent and child report. We note however, that while in general, parent ratings outperformed the child ratings, the mania/hypomania CAT was slightly more predictive of bipolar disorder for the child versus the parent ratings. It is interesting to examine the proportion of items adaptively administered from each of the subdomains for the parent and child. For the child, 41% of the items were mood, 47% behavior, 12% cognition and less than 1% rhythmicity. For the parent, 34% of the items were mood, 36% behavior, 30% cognition and less than 1% rhythmicity (there are only 11 rhythmicity items in the bank). Parents receive more cognitive items whereas the child receives more mood and behavior items.

We now place these findings in the context of the study's limitations, strengths, and the literature. This study's significant strengths included sample size, broad coverage of major psychiatric disorders in childhood, and use of a large item bank from which these K-CATs were derived. The study used of two independent samples, one for estimating the model

parameters of the MIRT model and developing the K-CAT and a second independent sample to demonstrate the validity of the K-CAT against semi-structured clinical interview-based diagnoses and extant scalar measures. In terms of the ROC curves for the prediction of the KSADS diagnoses from the K-CAT scores, we used three-fold cross-validation in which the same data were not used for the purpose of estimating the logistic regression and computing its predictive accuracy (i.e. ROC curve and corresponding AUC). Finally, we were also able to demonstrate test-retest reliability in this suite of K-CATs.

A limitation of the current study is that the participants are predominantly treatment-seeking psychiatric patients. The sample was not that diverse with respect to race or ethnicity and was restricted to English speakers. It remains to be seen how these measures will perform in other settings, such as pediatric primary care, schools, or pediatric emergency departments or in a more diverse sample. However, our adult CATs have been shown to perform well in non-psychiatric settings and with more diverse samples.^{32,33} This study excluded those with autistic spectrum, low intelligence, and psychosis, all conditions that may be encountered in routine practice. In addition to the above-noted exclusionary conditions, this study did not develop adaptive screens for common and impairing conditions such as post-traumatic stress disorder and alcohol and substance abuse. A third limitation is that the K-CAT time frame is 2 weeks with the exceptions of conduct disorder which is the past year. However, the goal of the K-CAT is to help the clinician efficiently focus on current psychopathology, which is more relevant for decision making, for the most part, than past history. Finally, although the K-CAT-SS accurately assesses the presence and severity of suicidal ideation, the concurrent and predictive validity with respect to suicidal behavior is unknown.

The K-CATs validated in this study make a unique contribution to the literature and provide tools that are not found elsewhere. The National Institutes of Health PROMIS program has developed adaptive screens for pediatric depression and anxiety.³⁴ A report examining the performance of these PROMIS adaptive measures compared to PROMIS static short forms concluded that the PROMIS CATs appeared to have “limited usefulness, over and above what can be accomplished by existing short static forms (8–10 items).”³⁵ In contrast to the PROMIS measures, the K-CATs: (1) were derived using multidimensional rather than unidimensional IRT, which allows for a greater number of items to be retained in the item bank; (2) use a more precise termination rule (i.e. 25% less uncertainty); (3) have been validated against structured diagnostic interviews and extant clinical measures; and (4) measured mental health constructs beyond just depression and anxiety. These features in the development process contributed to the high AUCs reported in this study between K-CATs and diagnoses.

Relative to other extant traditional fixed-length measures, the K-CAT is substantially shorter than the SCARED (41 items) and the MFQ (33 items). The SCARED was recently tested against a diagnostic interview with AUC's for parent and child report of 0.66 and 0.72, much lower than the K-CAT for anxiety.³⁶ Both the 13 and the 33 item MFQ have shown fairly high AUC's for both parent and child report (>0.83), but at the ideal cut-point, the sensitivity for parent and child report for the long MFQ is below 70%.^{37,38} The Strengths and Difficulties Questionnaire³⁹ is a very widely used 25-item parent report for general

psychopathology that is highly specific for “hyperactivity” and conduct disorder (>90%), but with lower sensitivity (68–74%, respectively).

In conclusion, these KCATs appear to be ready to be tested in variety of clinical settings. In specialty mental health care, these measures could help to conserve clinician assessment time. Moreover, these measures may be useful for the ongoing monitoring of treatment response and provide a foundation for the implementation of measurement-based care. These measures may also be useful in clinical research by facilitating the recruitment of large numbers of participants for phenotypic characterization and for pragmatic clinical trials.

These measures are also likely to be useful in non-psychiatric settings, such as emergency departments and pediatric primary care, where the main requirements are for the accurate and rapid assessment of acuity and disposition. The ability to identify patients who may need specialty care and/or hospitalization is predicated upon the ability to accurately assess diagnosis and gauge severity, which characterize these K-CAT measures. Consequently, the K-CAT could be used to develop clinician support tools that could help formulate recommendations that best fit patient needs. For example, adult versions of these measures are currently being deployed in a large public university in order to match students to the proper intensity and acuity of mental health care.

Future work will include prospective validation of the K-CAT-SS as a predictor of future suicidal behavior, deployment of the K-CATs in more diverse populations in emergency rooms and pediatric primary care settings and testing whether use of these measures can streamline assessments and support measurement-based care in specialty mental health settings.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The K-CAT™ will be licensed for research and clinical use by Adaptive Testing Technologies (www.adaptivetestingtechnologies.com). The study was supported by the National Institutes of Health (NIH) grant MH100155, with support for recruitment from the National Center for Advancing Translational Sciences (NCATS) grant TR001857.

Drs. Gibbons and Kim and Ms. Porta served as the statistical experts for this research.

Disclosure: Dr. Gibbons is a founder of Adaptive Testing Technologies, which distributes the CAT-MH™ battery of adaptive tests. The terms of this arrangement have been reviewed and approved by the University of Chicago in accordance with its conflict of interest policies. He has served as an expert witness for Merck, Pfizer, GlaxoSmithKline, and the US Department of Justice. Dr. Kupfer is a founder of Adaptive Testing Technologies, which distributes the CAT-MH™ battery of adaptive tests. He has an equity interest in Adaptive Testing Technologies, Inc. and in HealthRhythms, Inc. of which he is a founder. He is a board member and holds an equity interest in Minerva Neuroscience, has served on an Advisory Board for and received honoraria from Servier, and has received royalties for the Pittsburgh Sleep Quality Index (PSQI) from the University of Pittsburgh. Dr. Frank is a founder of Adaptive Testing Technologies, which distributes the CAT-MH™ battery of adaptive tests. She has an equity interest in Adaptive Testing Technologies, Inc. and in HealthRhythms, Inc. and is a founder and an employee of HealthRhythms, Inc. She has received royalties from the American Psychological Association Press, has received honoraria from Pfizer, and has served on an advisory board for and received honoraria from Servier. Dr. Brent has received research support from the National Institute of Mental Health, the American Foundation for Suicide

Prevention, the Once Upon a Time Foundation, and the Beckwith Foundation, has received royalties from Guilford Press from the electronic self-rated version of the C-SSRS from eRT, Inc. and from performing duties as an UptoDate Psychiatry Section Editor, and has received consulting fees from Healthwise. Drs. Lahey and Kim and Mss. George-Milford, Biernesser, Porta, and Moore report no biomedical financial interests or potential conflicts of interest.

References

1. Whitney DG, Peterson MD. US national and state-level prevalence of mental health disorders and disparities of mental health care use in children. *JAMA Pediatr.* 2019:E1–E3.
2. Merikangas KR, He JP, Burstein M, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication--Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry.* 2010;49:980–989. [PubMed: 20855043]
3. Simon AE, Pastor PN, Reuben CA, Huang LN, Goldstrom ID. Use of mental health services by children ages six to 11 with emotional or behavioral difficulties. *Psychiatr Serv.* 2015;66:930–937. [PubMed: 25975889]
4. Costello EJ, He JP, Sampson NA, Kessler RC, Merikangas KR. Services for adolescents with psychiatric disorders: 12-month data from the National Comorbidity Survey-Adolescent. *Psychiatr Serv.* 2014;65:359–366. [PubMed: 24233052]
5. Leverich GS, Post RM, Keck PE Jr., et al. The poor prognosis of childhood-onset bipolar disorder. *J Pediatr.* 2007;150:485–490. [PubMed: 17452221]
6. Perlis RH, Dennehy EB, Miklowitz DJ, et al. Retrospective age at onset of bipolar disorder and outcome during two-year follow-up: results from the STEP-BD study. *Bipolar Disord.* 2009;11:391–400. [PubMed: 19500092]
7. Wang PS, Berglund P, Olfson M, Pincus HA, Wells KB, Kessler RC. Failure and delay in initial treatment contact after first onset of mental disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62:603–613. [PubMed: 15939838]
8. Obradovic J, Burt KB, Masten AS. Testing a dual cascade model linking competence and symptoms over 20 years from childhood to adulthood. *J Clin Child Adolesc Psychol.* 2010;39:90–102. [PubMed: 20390801]
9. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry.* 2005;62:593–602. [PubMed: 15939837]
10. Kim-Cohen J, Caspi A, Moffitt TE, Harrington H, Milne BJ, Poulton R. Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Arch Gen Psychiatry.* 2003;60:709–717. [PubMed: 12860775]
11. Ng MY, Weisz JR: Annual Research Review: Building a science of personalized intervention for youth mental health. *J Child Psychol Psychiatry.* 2016 3;57(3):216–36. [PubMed: 26467325]
12. Nathan PI, Gorman JM (eds.) *Treatments That Work*, Oxford University Press, 2015
13. Gibbons RD, Weiss DJ, Frank E, Kupfer D. Computerized adaptive diagnosis and testing of mental health disorders. *Annu Rev Clin Psychol.* 2016;12:83–104. [PubMed: 26651865]
14. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of a computerized adaptive test for depression. *Arch Gen Psychiatry.* 2012;69:1104–1112. [PubMed: 23117634]
15. Gibbons RD, Weiss DJ, Pilkonis PA, et al. Development of the CAT-ANX: a computerized adaptive test for anxiety. *Am J Psychiatry.* 2014;171:187–194. [PubMed: 23929270]
16. Gibbons RD, Kupfer D, Frank E, Moore T, Beiser DG, Boudreaux ED. Development of a Computerized Adaptive Test Suicide Scale-The CAT-SS. *J Clin Psychiatry.* 2017;78:1376–1382. [PubMed: 28493655]
17. Kaufman J, Birmaher B, Brent D, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry.* 1997;36:980–988. [PubMed: 9204677]
18. 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;168:1266–1277. [PubMed: 22193671]

19. Angold A, Costello EJ, Messer SC, Pickles A, Winder F, & Silver D (1995) The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *International Journal of Methods in Psychiatric Research*, 5, 237–249.
20. Birmaher B, Khetarpal S, Brent D, et al.: The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. *J Am Acad Child Adolesc Psychiatry*. 1997;36:545–553 [PubMed: 9100430]
21. Henry David B.; Pavuluri Mani N.; Youngstrom Eric; Birmaher Boris (2008-4-01). “Accuracy of brief and full forms of the child mania rating scale” (PDF). *Journal of Clinical Psychology*. 64 (4): 368–381. [PubMed: 18302291]
22. Shaffer D, Gould MS, Brasic J, et al. A children’s global assessment scale (CGAS). *Arch Gen Psychiatry*. 1983;40:1228–1231. [PubMed: 6639293]
23. Gardner W, Lucas A, Kolko DJ, Campo JV. Comparison of the PSC-17 and alternative mental health screens in an at-risk primary care sample. *J Am Acad Child Adolesc Psychiatry*. 2007;46:611–618. [PubMed: 17450052]
24. Lahey BB, Applegate B, Waldman ID, Loft JD, Hankin BL, Rick J. The structure of child and adolescent psychopathology: generating new hypotheses. *J Abnorm Psychol*. 2004;113:358–385. [PubMed: 15311983]
25. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Washington, DC: American Psychiatric Association; 2013.
26. Kincaid JP, Fishburne RP Jr, Rogers RL, Chissom BS. Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel. Institute for Simulation and Training Paper 56 1975.
27. Gibbons RD, Hedeker DR. Full-information item bi-factor analysis. *Psychometrika*. 1992;57:423–436.
28. Gibbons RD, Bock RD, Hedeker D, et al. Full-information item bifactor analysis of graded response data. *Appl Psychol Meas*. 2007;31:4–19.
29. Holzinger KJ, Swineford F. The bi-factor method. *Psychometrika*. 1937;2:41–54.
30. Jiang S, Wang C, Weiss DJ. Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in psychology*. 2016; 7:109 [PubMed: 26903916]
31. De Los Reyes A, Augenstein TM, Wang M, et al. The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological bulletin* 2015; 141(4): 858. [PubMed: 25915035]
32. Graham AK, Minc A, Staab E, Beiser DG, Gibbons RD, Laiteerapong N. Validation of the Computerized Adaptive Test for Mental Health in Primary Care. *Ann Fam Med*. 2019;17:23–30. [PubMed: 30670391]
33. Gibbons RD, Alegria M, Cai L, et al. Successful validation of the CAT-MH Scales in a sample of Latin American migrants in the United States and Spain. *Psychol Assess*. 2018;30:1267–1276. [PubMed: 29792502]
34. Irwin DE, Stucky B, Langer MM, et al. An item response analysis of the pediatric PROMIS anxiety and depressive symptoms scales. *Qual Life Res*. 2010;19:595–607. [PubMed: 20213516]
35. Varni JW, Magnus B, Stucky BD, et al. Psychometric properties of the PROMIS (R) pediatric scales: precision, stability, and comparison of different scoring and administration options. *Qual Life Res*. 2014;23:1233–1243. [PubMed: 24085345]
36. Ivarsson T, Skarphedinsson G, Andersson M, et al. The Validity of the Screen for Child Anxiety Related Emotional Disorders Revised (SCARED-R) Scale and Sub-Scales in Swedish Youth. *Child Psychiatry Hum Dev*. 2018; 49(2): 234–243. [PubMed: 28756556]
37. Daviss WB, Birmaher B, Melhem NA, et al. Criterion validity of the Mood and Feelings Questionnaire for depressive episodes in clinic and non-clinic subjects. *J Child Psychol Psychiatry*. 2006; 47(9): 927–34. [PubMed: 16930387]
38. Thapar A, McGuffin P. Validity of the shortened Mood and Feelings Questionnaire in a community sample of children and adolescents: a preliminary research note. *Psychiatry Res*. 1998; 81(2): 259–68. [PubMed: 9858042]

39. Goodman R Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry*. 2001 40(11): 1337–45. [PubMed: 11699809]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Clinical Vignette

Although further work is required, we believe that our computerized adaptive tests could be utilized in the future in the following manner:

A 14-year old boy is brought for an assessment for irritability, impulsivity, mood lability, and inattention. The patient and his mother first filled out computerized adaptive tests (CATs) for depression, anxiety, attention deficit disorder, hyperactivity (ADHD), mania, oppositional-defiant disorder, conduct disorder, and suicidal risk, taking the parent and child less than 10 minutes each. Using a combination of information from the patient and parent, these CATs indicated with a high degree of likelihood, the presence of mania, depression, attention-deficit disorder, hyperactivity (ADHD), and suicidal risk. With guidance from the results of these CATs, the assessing clinician determined the age of onset, course, past attempts at treatment, and contribution to functional impairment from these different conditions. The patient was diagnosed with major depressive disorder, bipolar disorder, NOS, ADHD, and manifested suicidal ideation with intent but without a plan. Subsequent exploration of self-harm indicated that the patient was engaged in non-suicidal self-injury several times per week. Because the patient had evidence of a bipolar spectrum disorder, he was not started on an antidepressant, but instead was treated with a mood stabilizer. The patient was able to develop a safety plan with the assessing clinician, but, given the combination of suicidal ideation with intent and frequent non-suicidal self-injury, the patient was started in an intensive outpatient program that emphasized dialectic behavioral skills. The patient's ADHD was not targeted because it was the summer and his major difficulties from this condition were manifested in school.

The patient's clinical status was evaluated weekly using the relevant CATs (depression, mania, ADHD, suicidality), and based on continued depressive symptoms, an antidepressant was initiated. With improvement in manic, depressive, and suicidal symptoms over the next several weeks, the patient was successfully transitioned to weekly outpatient therapy, with relief of his non-suicidal self-injury.

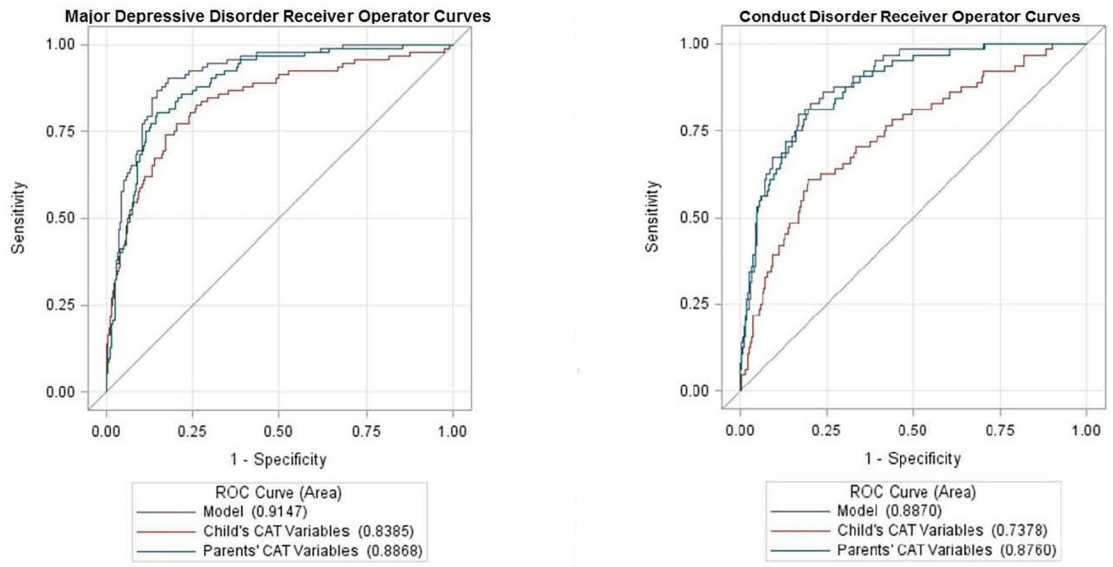


Figure 1: Receiver Operator Curves (ROC) for Diagnosis of Major Depressive Disorder and Conduct Disorder Based on Child and Parent Computerized Adaptive Tests (Model = Combined)
Note: CAT = Computerized Adaptive Test.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

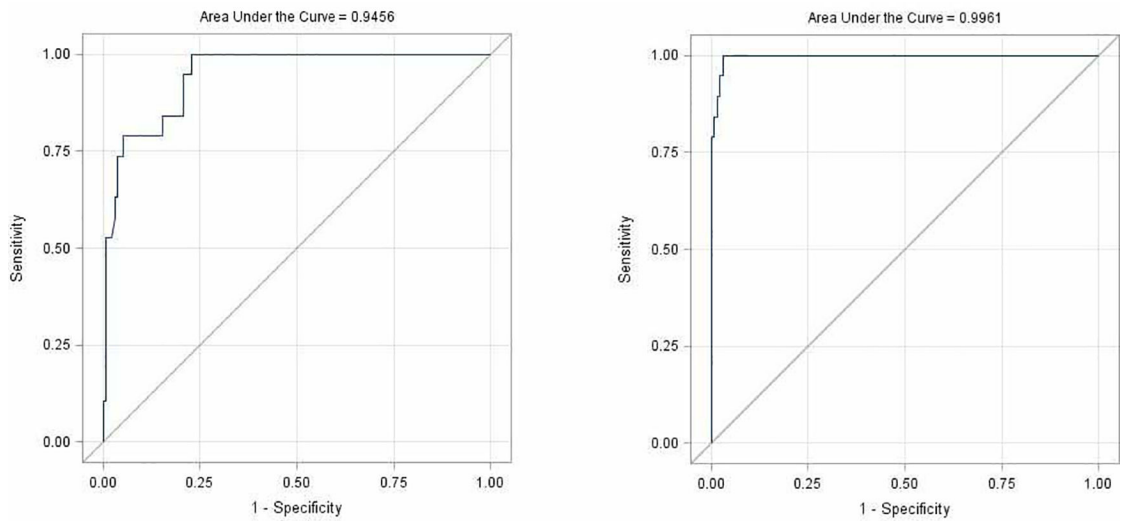


Figure 2: Receiver Operator Curves for Columbia-Suicide Severity Rating Scale Suicidal Ideation Using the Kiddie-Computerized Adaptive Test –Suicide Scale Alone and the Kiddie-Computerized Adaptive Test for Parent and Child and the Kiddie-Computerized Adaptive Test –Suicide Scale for the Child

Table 1.

Demographic Characteristics of the Sample

	MDD	Anxiety	Mania	ADHD	OCD	CD	Control	Total
N	137	124	68	115	103	54	200	801
Age, M (SD)	14.7 (1.9)	13.8 (2.6)	13.3 (2.8)	11.1 (3.2)	11.1 (3.0)	12.4 (3.0)	12.2 (3.1)	12.7 (3.1)
Sex, n (% Female participants)	101 (73.7)	88 (71.0)	43 (63.2)	37 (32.2)	42 (40.8)	25 (46.3)	108 (54.0)	444 (55.4)
Race, n (% Caucasian)	104 (75.9)	107 (86.3)	53 (77.9)	41 (35.7)	37 (35.9)	22 (40.7)	126 (63.0)	490 (61.2)
Hispanic, n (% Yes)	14 (10.9)	4 (3.5)	5 (7.9)	7 (6.3)	5 (5.2)	1 (2.0)	5 (2.6)	41 (5.4)
Family Income, n (% Yes)								
<\$24,999	33 (24.3)	23 (19.0)	15 (22.1)	49 (43.4)	61 (60.4)	36 (66.7)	38 (19.0)	255 (32.2)
\$25,000-\$49,999	28 (20.6)	25 (20.7)	27 (39.7)	33 (29.2)	20 (19.8)	9 (16.7)	45 (22.5)	187 (23.6)
\$50,000-\$74,999	19 (14.0)	18 (14.9)	12 (17.7)	12 (10.6)	11 (10.9)	4 (7.4)	38 (19.0)	114 (14.4)
\$75,000-\$99,999	22 (16.2)	13 (10.7)	5 (7.4)	7 (6.2)	2 (2.0)	2 (3.7)	32 (16.0)	83 (10.5)
\$100,000	34 (25.0)	42 (34.7)	9 (13.2)	12 (10.6)	7 (6.9)	3 (5.6)	47 (23.5)	154 (19.4)
Grade, M (SD)	9.2 (2.0)	8.4 (2.7)	7.6 (3.5)	5.5 (3.2)	5.6 (3.0)	6.8 (3.2)	6.5 (3.2)	7.1 (3.2)

Note: ADHD = attention-deficit/hyperactivity disorder; CD = conduct disorder; MDD = major depressive disorder; OCD = obsessive-compulsive disorder.

Table 2:

Odds Ratios and 95% Confidence Intervals for Association Between Kiddie Schedule for Affective Disorders and Schizophrenia Diagnoses and a 10 Point Difference in Individual Kiddie-Computerized Adaptive Test Measures

Diagnosis/K-CAT Measure	Child	Parent	Max of Child and Parent
MDD/K-CAT-Depression	2.100 (1.757, 2.524)	2.886 (2.282, 3.643)	3.162 (2.456, 4.046)
GAD/K-CAT-Anxiety	2.179 (1.825, 2.524)	1.459 (1.280, 1.660)	1.913 (1.614, 2.261)
ADHD/K-CAT-ADHD	1.280 (1.172, 1.397)	2.220 (1.949, 2.547)	2.119 (1.842, 2.433)
CD/K-CAT-CD	1.384 (1.219, 1.568)	2.061 (1.724, 2.478)	1.931 (1.629, 2.303)
ODD/K-CAT-ODD	1.466 (1.318, 1.629)	2.282 (1.967, 2.665)	2.501 (2.119, 2.944)
BP1 and BP2/K-CAT-Mania	1.424 (1.094, 1.860)	1.293 (1.030, 1.629)	1.480 (1.138, 1.931)

Note: ADHD = attention-deficit/hyperactivity disorder; BP1 = Bipolar 1; BP2 = Bipolar 2; CD = conduct disorder; GAD = generalized anxiety disorder; K-CAT = Kiddie-Computerized Adaptive Test; MDD = major depressive disorder; ODD = oppositional defiant disorder.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Area Under the Curve (AUC) and 95% Confidence Intervals for Association Between Kiddie Schedule for Affective Disorders and Schizophrenia Diagnoses and Combined Kiddie-Computerized Adaptive Test Measures

Diagnosis	Child	Parent	Child and Parent
MDD	.839 (.790, .887)	.887 (.853, .921)	.915 (.886, .943)
GAD	.806 (.764, .848)	.752 (.708, .797)	.829 (.790, .867)
ADHD	.712 (.671, .753)	.845 (.814, .876)	.858 (.829, .888)
CD	.738 (.671, .805)	.876 (.835, .917)	.887 (.850, .925)
ODD	.758 (.715, .796)	.860 (.830, .890)	.877 (.849, .904)
BP1 or BP2	.737 (.621, .853)	.738 (.630, .846)	.847 (.788, .906)

Note: ADHD = attention-deficit/hyperactivity disorder; BP1 = Bipolar 1; BP2 = Bipolar 2; CD = conduct disorder; GAD = generalized anxiety disorder; MDD = major depressive disorder; ODD = oppositional defiant disorder.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript