Original Research Article

Assessment
1–13

DOI: 10.1177/10731911221143994
journals.sagepub.com/home/asm
SAGE

The DIAMOND-KID: Psychometric Properties of a Structured Diagnostic Interview for *DSM-5* Anxiety, Mood, and Obsessive-Compulsive and Related Disorders in Children and Adolescents

David F. Tolin[1,2], Kimberly S. Sain[1], Elizabeth Davis[1], Christina Gilliam[1], Scott E. Hannan[1], Kristen S. Springer[1], Jessica Stubbing[1], Jamilah R. George[1,3], Aniskha Jean[1], Rachel Goldblum[1], Benjamin W. Katz[1], Kate Everhardt[1], Sabrina Darrow[4], Elizabeth Ellis Ohr[5], Matthew E. Young[6], and Marisa D. Serchuk[7]

Abstract

The objective of the present study was to examine the reliability and validity of a new semi-structured interview for pediatric psychiatric disorders, which is needed as existing interviews do not cover the full range of anxiety, mood, and obsessive-compulsive disorder (OCD)-related disorders. Three hundred eleven child patients (aged 10–17) were administered the Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders-Child and Adolescent Version (DIAMOND-KID). Of these, 65 provided interrater reliability data and 59 provided test–retest reliability data. Participants also completed self-report measures that assessed symptoms of anxiety, mood, and OCD and related disorders. Although parents/guardians could participate in the interview at the clinician's discretion, most of the initial interviews and all of the reliability interviews were based on the child's self-report. Test–retest reliability ranged from very good to excellent. Interrater reliability was more variable, with estimates for generalized anxiety disorder and major depressive disorder in the questionable range; the other interrater reliability estimates ranged from good to very good. Convergent validity was established by significant between-group comparisons on applicable self-report measures for all diagnoses. The results of the present study indicate that the DIAMOND-KID is a promising semi-structured diagnostic interview for 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* in pediatric populations.

Keywords

anxiety disorders, mood disorders, obsessive-compulsive and related disorders, interview, diagnosis

The *Diagnostic Interview for Anxiety, Mood, and Obsessive-Compulsive and Related Neuropsychiatric Disorders* (DIAMOND; Tolin et al., 2018) is a semi-structured interview for adults that corresponds to the diagnostic criteria specified in the 5th edition of the *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; *DSM-5*; American Psychiatric Association, 2013). While the DIAMOND assesses a broad range of psychiatric conditions, particular attention is paid to the anxiety, mood, and obsessive-compulsive disorder (OCD) and related diagnoses, with information about differential diagnosis, associated features, and specifiers provided to assist diagnosticians. Psychometric testing of the DIAMOND demonstrates very good to excellent interrater reliability and good to excellent test–retest reliability. Furthermore, the convergent validity of the DIAMOND is established by significant between-group comparisons on applicable self-report measures for nearly all diagnoses (Tolin et al., 2018).

[1]Institute of Living, Hartford, CT, USA
[2]Yale School of Medicine, New Haven, CT, USA
[3]University of Connecticut, Storrs, USA
[4]University of California, San Francisco, USA
[5]Private Practice, Portsmouth, NH, USA
[6]The University of Chicago Pritzker School of Medicine, IL, USA
[7]Illinois Institute of Technology, Chicago, USA

Corresponding Author:
David F. Tolin, The Institute of Living, 200 Retreat Avenue, Hartford, CT 06106, USA.
Email: david.tolin@hhchealth.org

The present article describes the development and testing of a pediatric version, the *Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders: Child and Adolescent Version* (DIAMOND-KID). There is a clear need for well-validated diagnostic measures for children and adolescents. When conducting clinical research trials, researchers must be able to define their sample adequately and determine whether study participants meet the criteria for inclusion or exclusion diagnoses. In clinical settings, clinicians are often faced with challenging differential diagnostic cases and must be able to evaluate patients according to clearly defined diagnostic criteria. Finally, in training programs, structured diagnostic interviews are frequently used to teach the process of diagnostic interviewing and to familiarize trainees with diagnostic criteria.

For a review of existing diagnostic interviews for children and adolescents, the reader is directed to Leffler et al. (2015). Because of the substantial changes in diagnostic categories and criteria (including new diagnoses) from *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; *DSM-IV*; American Psychiatric Association, 2000) to *DSM-5*, we will focus here on those interviews that are specifically designed to assess *DSM-5* diagnoses. These include the *Child and Adolescent Psychiatric Assessment* (CAPA; Angold et al., 1995), the *Schedule for Affective Disorders and Schizophrenia for School-Aged Children* (K-SADS; Kaufman et al., 1997), the *Children's Interview for Psychiatric Symptoms* (ChIPS; Young et al., 2016), the *Mini-International Neuropsychiatric Interview for Children and Adolescents* (MINI-KID; Sheehan et al., 2010), and the *Development and Wellbeing Assessment* (DAWBA; Aebi et al., 2012). As shown in Table 1, none of these interviews measures anxiety, mood, and OCD and related disorders in their entirety. In particular, OCD and related disorders are not well represented in existing interviews. This is not a trivial oversight. In a large sample of the adolescent general population, 4% reported clinically significant symptoms of body dysmorphic disorder, 1% of hoarding disorder, 1% of trichotillomania, and 2% of skin picking disorder (Moreno-Amador et al., 2022). The rate is likely higher among clinically referred samples: nearly 7% of adolescent inpatients may have body dysmorphic disorder (Dyl et al., 2006); hoarding behaviors are noted in more than one-quarter of children with OCD (Hojgaard et al., 2019); and trichotillomania is present in nearly 10% of children referred for evaluation of alopecia (Stroud, 1983). We suggest, therefore, that these conditions are not rare, particularly in clinical populations. Thus, there is a need for a diagnostic interview that assesses the full range of *DSM-5* disorders.

Evaluation of the DIAMOND-KID may also help shed light on the reliability of the *DSM-5* diagnoses themselves, which is unclear in children. The *DSM-5* field trials included child and adolescent participants (Clarke et al., 2013). However, few of the anxiety, mood, and OCD and related disorders were examined. Only mixed anxiety depressive disorder was assessed within the anxiety disorders, showing an unacceptable test–retest reliability of $\kappa = 0.05$. Within the mood disorders, only major depressive disorder (MDD) and disruptive mood dysregulation disorder (DMDD) were examined; these diagnoses showed questionable test–retest reliabilities of $\kappa = 0.28$ and 0.25, respectively (Regier et al., 2013). No OCD and related disorders were assessed.

Among adult participants in the *DSM-5* field trials, within the anxiety disorders, generalized anxiety disorder (GAD) showed a questionable test–retest reliability of $\kappa = 0.20$. Within the mood disorders, test–retest reliability for bipolar I disorder was good ($\kappa = 0.56$), although the reliability for MDD was questionable ($\kappa = 0.28$). The investigators were unable to obtain accurate estimates of κ (defined as a standard error of less than or equal to 0.1 and a 95% confidence interval of less than or equal to 0.5) for bipolar II disorder or hoarding disorder (HD; Regier et al., 2013).

We note that many have questioned the validity of syndromal models of psychopathology in general and of the *DSM-5* in particular (e.g., Brown & Barlow, 2009; Cuthbert, 2014; Kendell & Jablensky, 2003; Kotov et al., 2017; Krueger & Bezdjian, 2009; Krueger et al., 2005). To the extent that the "validity" of a *DSM-5* interview can be examined, such examination is limited at present to understanding the interview's fidelity to the syndromal model implied by the *DSM-5*, rather than the validity of the model itself. Nevertheless, despite its limitations, the *DSM-5* remains the most commonly used diagnostic system for research, treatment, and clinical training in the United States (Tyrer, 2014), and the categorical model of psychopathology remains the dominant scheme in health care (First, 2005). Thus, empirically validated semi-structured interviews that correspond to the *DSM-5* are needed.

The aim of the present study was to examine the reliability and validity of the DIAMOND-KID in a treatment-seeking sample of children and adolescents. It was predicted that the DIAMOND-KID would show adequate interrater reliability for anxiety, mood, and OCD and related disorders. Although the checklist used in the *DSM-5* field trials yielded generally unacceptable reliability coefficients (Regier et al., 2013), we predicted that the more detailed investigation of specific symptoms in the DIAMOND-KID would improve reliability. It was further predicted that the measure would show adequate test–retest reliability and convergent validity, as measured by dimensional self-report measures, for those disorders. In the following section, we report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

Table 1. Coverage of the DSM-5 Anxiety, Mood, and Obsessive-Compulsive and Related Disorders by Existing Diagnostic Interviews.

Diagnosis	CAPA	K-SADS	ChIPS	MINI-Kid	DAWBA	DISK	DIAMOND-KID
Anxiety disorders							
Selective mutism	+	+	−	−	−	+	+
Social anxiety disorder/social phobia	+	+	+	+	+	+	+
Panic disorder	+	+	−	+	+	+	+
Agoraphobia	+	+	−	+	+	+	+
Generalized anxiety disorder	+	+	+	+	+	+	+
Specific phobia	+	+	+	+	+	+	+
Separation anxiety disorder	+	+	+	+	+	+	+
Mood disorders							
Persistent depressive disorder	+	+	+	−	−	+	+
Major depressive disorder	+	+	+	+	+	+	+
Bipolar I disorder	−	+	+	+	−	+	+
Bipolar II disorder	−	+	+	+	−	+	+
Cyclothymic disorder	−	+	−	−	−	−	+
Disruptive mood dysregulation disorder	−	+	−	−	−	+	+
Premenstrual dysphoric disorder	−	−	−	−	−	−	+
Obsessive-compulsive and related disorders							
Obsessive-compulsive disorder	+	+	+	+	+	+	+
Body dysmorphic disorder	−	−	−	−	−	−	+
Hoarding disorder	−	−	−	−	−	−	+
Trichotillomania	+	−	−	−	−	+	+
Excoriation (skin-picking) disorder	−	−	−	−	−	+	+

Note. CAPA = Child and Adolescent Psychiatric Assessment; K-SADS = Schedule for Affective Disorders and Schizophrenia for School-Aged Children; ChIPS = Children's Interview for Psychiatric Symptoms; MINI-KID = Mini-International Neuropsychiatric Interview for Children and Adolescents; DAWBA = Development and Wellbeing Assessment; DISK = Diagnostic Interview Schedule for Children; DIAMOND-KID = Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders: Child and Adolescent Version; OCD = obsessive-compulsive disorder.

Method

Participants

The DIAMOND-KID was administered as part of a routine intake to consecutive child and adolescent (ages 6–17) English-speaking patients who were seeking treatment at a hospital-based outpatient clinic for anxiety, mood, and obsessive-compulsive and related disorders; two clinical research labs specializing in pediatric psychopathology; or a private practice. Because the number of children younger than the age of 10 was small ($n = 19$), data from children in that age range will not be analyzed here, and the sample is age 10 to 17. It is therefore noted that the results presented here may not apply to younger children.

Following the initial interview, participants were then invited to participate in the interrater and test–retest reliability arm of the study. No participants were excluded due to comorbid psychiatric conditions, severity of illness, clarity of diagnosis, or complexity of illness.

To calculate the sample size, we used an online calculator (Arifin, 2022). With an expected $\kappa = 0.4$ and a null hypothesis of $\kappa = 0.0$, assuming a 10% prevalence of any given disorder (as has been our experience for most of the core DIAMOND-KID disorders in our clinical practice), a two-tailed test would have 80% power to detect κ with 50 participants. As shown in Table 2, 311 individuals aged 10 to 17 received the DIAMOND-KID. Of these, 65 provided interrater reliability data and 59 provided test–retest reliability data. Mean age of each sample was approximately 14 years. Skewness of age was -0.70 for the total sample, −0.35 for the interrater reliability sample, and −0.44 for the test–retest sample, suggesting that these distributions did not differ significantly from normal. The samples included more girls than boys. Of the diagnoses assigned during the initial DIAMOND administration, the most common anxiety disorders were social phobia and GAD. The most common mood disorder was MDD. The most common obsessive-compulsive and related disorder was OCD. Of note, some anxiety, mood, and obsessive-compulsive and related disorders were under-represented in the present study, particularly selective mutism (SM), bipolar disorders, DMDD, premenstrual dysphoric disorder, body dysmorphic disorder (BDD), and HD. We opted to analyze

Table 2. Sample Description.

	Total sample ($N = 311$)	Interrater reliability subsample ($n = 65$)	Test-retest reliability subsample ($n = 59$)
Demographics			
Age, *M* (*SD*)	14.33 (2.02)	14.46 (2.19)	14.63 (2.18)
Female, *N* (%)	197 (64.8%)	48 (73.8%)	44 (74.6%)
White, *N* (%)	258 (84.6%)	50 (76.9%)	46 (78.0%)
Black, *N* (%)	9 (3.0%)	4 (6.2%)	4 (6.8%)
Hispanic, *N* (%)	15 (4.9%)	4 (6.2%)	3 (5.1%)
Intake diagnoses			
Anxiety disorders			
SM *N* (%)	5 (1.6%)	1 (1.5%)	1 (1.7%)
SoP *N* (%)	114 (36.7%)	22 (33.8%)	21 (35.6%)
PD *N* (%)	24 (7.7%)	2 (3.1%)	2 (3.4%)
AGO *N* (%)	18 (5.8%)	3 (4.6%)	3 (5.1%)
GAD *N* (%)	97 (31.2%)	24 (36.9%)	23 (39.0%)
SpP *N* (%)	38 (12.2%)	11 (16.9%)	12 (20.3%)
SAD *N* (%)	9 (2.9%)	2 (3.1%)	1 (1.7%)
Any *N* (%)	203 (65.3%)	43 (66.2%)	41 (69.5%)
Bipolar disorders			
BP1 *N* (%)	1 (0.3%)	1 (1.5%)	1 (1.7%)
BP2 *N* (%)	3 (1.0%)	2 (3.1%)	2 (3.4%)
CYC *N* (%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
Any *N* (%)	4 (1.3%)	3 (4.6%)	3 (5.1%)
Depressive disorders			
PDD *N* (%)	40 (12.9%)	8 (12.3%)	8 (13.6%)
MDD *N* (%)	78 (25.1%)	18 (27.7%)	17 (28.8%)
DMDD *N* (%)	2 (0.6%)	1 (1.5%)	1 (1.1%)
PMDD *N* (%)	5 (1.6%)	2 (3.1%)	2 (3.4%)
Any *N* (%)	116 (37.3%)	26 (40.0%)	25 (42.4%)
Obsessive-compulsive and related disorders			
OCD *N* (%)	127 (40.8%)	26 (40.0%)	22 (37.3%)
BDD *N* (%)	7 (2.3%)	2 (3.1%)	2 (3.4%)
HD *N* (%)	1 (0.3%)	0 (0.0%)	0 (0.0%)
TTM *N* (%)	21 (6.8%)	5 (7.7%)	5 (8.5%)
EXD *N* (%)	26 (8.4%)	9 (13.8%)	10 (16.9%)
Any *N* (%)	161 (51.8%)	36 (53.7%)	31 (52.5%)
Other disorders			
PTSD *N* (%)	16 (5.1%)	3 (4.6%)	3 (5.1%)
IAD *N* (%)	3 (1.0%)	0 (0.0%)	0 (0.0%)
SUD *N* (%)	3 (1.0%)	1 (1.5%)	1 (1.7%)
ADHD *N* (%)	34 (10.9%)	14 (21.5%)	14 (23.7%)
ODD *N* (%)	10 (3.2%)	1 (1.5%)	1 (1.7%)
Any ED *N* (%)	8 (2.6%)	5 (7.5%)	4 (6.8%)
Any Tic *N* (%)	13 (4.2%)	4 (6.0%)	4 (6.8%)

Note. SM = selective mutism; SoP = social anxiety disorder (social phobia); PD = panic disorder; AGO = agoraphobia; GAD = generalized anxiety disorder; SpP = specific phobia; SAD = separation anxiety disorder; BP1 = bipolar I disorder; BP2 = bipolar II disorder; CYC cyclothymia; PDD = persistent depressive disorder (dysthymia); MDD = major depressive disorder; DMDD = disruptive mood dysregulation disorder; PMDD = premenstrual dysphoric disorder; OCD = obsessive-compulsive disorder; BDD = body dysmorphic disorder; HD = hoarding disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; PTSD = posttraumatic stress disorder; IAD = illness anxiety disorder; SUD = substance use disorder; ADHD = attention-deficit/hyperactivity disorder; ODD = oppositional defiant disorder; ED = eating disorder.

only those diagnoses represented in eight or more children. Therefore, reliability and validity analyses were not conducted on these diagnoses. For sample description purposes, Table 2 also shows rates of posttraumatic stress disorder (PTSD), illness anxiety disorder, substance use disorder (SUD), attention-deficit/hyperactivity disorder

(ADHD), oppositional defiant disorder, eating disorders, and tic disorders. Of these, only ADHD had enough diagnosed participants with self-report measures and repeat interviews for further analysis.

Measures

Diagnostic Interview for Anxiety, Mood, and Obsessive-Compulsive and Related Neuropsychiatric Disorders: Child and Adolescent Version. The DIAMOND-KID is a structured clinical interview that queries the *DSM-5* diagnostic criteria for anxiety disorders, bipolar disorders, depressive disorders, OCD and related disorders, trauma- and stressor-related disorders, schizophrenia spectrum and other psychotic disorders, feeding and eating disorders, somatic symptom and related disorders, substance-related and addictive disorders, and neurodevelopmental disorders. The primary focus of the DIAMOND-KID is the anxiety, mood, and obsessive-compulsive and related disorders, and the diagnostic criteria for these disorders are supplemented with more clinically relevant questions, such as symptom dimensions, as well as information about common differential diagnoses. The other disorders were included in the DIAMOND-KID because an accurate diagnosis of anxiety, mood, or OCD and related disorder requires the interviewer to carefully rule out alternative diagnoses (e.g., the diagnosis of BDD may require ascertaining that the patient's symptoms are not limited to weight or shape concerns secondary to an eating disorder; thus, effective BDD assessment requires assessment of certain eating disorders). A suicide screen is also included that queries suicidal ideation, intent, plan, means, behaviors, and protective factors. We also added screening modules for school refusal, autism spectrum disorder, and pediatric acute-onset neuropsychiatric syndrome (PANS). As noted earlier, the DIAMOND-KID can be administered to the child alone or to the child with a parent/guardian present.

The DIAMOND-KID was derived from the adult DIAMOND (Tolin et al., 2018). First, a panel of expert clinicians reviewed the DIAMOND items and suggested typical and atypical symptom presentations, initial questions, and follow-up questions. We then conducted initial feasibility testing, with iterative feedback from users at various stages of professional development. Specific changes from the DIAMOND to the DIAMOND-KID include the use of developmentally appropriate language (created with the input of child and adolescent clinical psychologists), the inclusion of a screening tool and interview questions for parents/guardians, the inclusion of child-specific diagnoses (e.g., DMDD), and the capacity for obtaining responses from parents or guardians (which can be helpful with younger children or with certain psychiatric disorders). We also included syndrome screens for school refusal, autism spectrum disorder, and PANS, given the frequency with which these problems are noted among children with anxiety, mood, and obsessive-compulsive and related disorders. Although these syndrome screens are not intended to be full evaluations, they were included to assist with differential diagnosis and to signal assessors when additional follow-up assessment might be required. Finally, the DIAMOND-KID was designed to be administered to the child alone, or to the child with a parent/guardian present; guidance is provided as to the best source of information (e.g., the diagnosis of obsessive-compulsive disorder can be assessed with the child with or without a parent/guardian, whereas the diagnosis of the oppositional defiant disorder is best assessed with the parent with input from the child where appropriate).

The wording of questions was altered from the *DSM-5* criteria, and the questions were ordered to minimize clinician and patient burden. Questions were directed toward the child, unless the parent or guardian was present in which case either the child or the parent/guardian was allowed to respond. Specifically, the following format was used:

1. Initial questions: Preliminary questions allow the interviewer to obtain an overview of the presenting problem prior to inquiring about specific diagnoses. These include:
    a. "Can you describe what kind of problem or problems you are here to discuss?"
    b. "How is your physical health? Do you have any significant medical conditions?"
    c. "What medications do you currently take?"
    d. "Have you had mental health treatment before? If so, can you describe it? When did it occur?"
    e. "Have you ever been hospitalized for psychiatric reasons before? If so, can you describe it? Where and when were you hospitalized?"
    f. "Does anyone in your family have a history of mental health problems? What kind of problems?"
    g. "Have you been having any thoughts about hurting or killing yourself?"
    h. "How is your sleeping? Is it hard for you to fall asleep? Do you wake up a lot during the night?"
    i. "How is school going? What kind of grades do you get? Do you ever get in trouble at school? Are you in any special programs at school or receive any special accommodations or special education?"
    j. "How are things going at home? How well do you get along with your parents and siblings? Do you have a lot of arguments? Do you ever get in trouble at home?"
    k. "How are things going with friends? Do you have many friends? How often do you do things with friends outside of school? When you get

together with friends, is it in person or over the Internet?"

l. [For parent/guardian]: "How was your child's early development? Were there any delays or concerns? What was his or her general temperament like (cheerful, fussy, shy, etc.)?"

m. "Are there aspects of your background or identity that impact [problem described], or that are relevant for me to know? By background or identity, I mean, for example, the communities you belong to, the languages you speak, where you or your family are from, your race or ethnic background, your gender identity or sexual orientation, or your faith or religion."

2. Symptom questions: The specific *DSM-5* symptoms were listed and queried (self-report and parent-report screening forms, consisting of the lead symptom questions [e.g., in the case of OCD, one question about the presence of obsessions and one about the presence of compulsions], were developed to facilitate this process, although interviewers were allowed to probe any symptoms mentioned during the interview, regardless of the participant's response on the screening form). Symptoms (coded as present or absent) were queried over the past month, except when a longer duration is needed for diagnostic purposes (e.g., a lifetime history of manic episodes for a bipolar disorder diagnosis). For example, in the case of OCD, the interviewer asked the following questions, required to satisfy Criterion A1, "Recurrent and persistent thoughts, urges, or images that are experienced, at some time during the disturbance, as intrusive and unwanted, and that in most individuals cause marked anxiety or distress":
   a. "In the past month, do you have a lot of thoughts, worries, or images in your mind that you don't want to have? Some examples are thoughts that you will get dirt or germs on you, that you will make a terrible mistake, or being very uncomfortable if things aren't just right."
3. Clarifying questions: For additional clarity, when appropriate the interviewer asked questions about specific symptom dimensions. For example, after the aforementioned symptom question for OCD, the interviewer asked:
   a. "What kind of thoughts do you have?"
   b. "Do these thoughts come into your mind even when you don't want them to?"
   c. "Do they come into your mind again and again and bother you for some time?"
4. A checklist was then provided to the interviewer, allowing the interviewer to check dimensions such as fear of contamination, fear of harming others, worrying about others being hurt, need for evenness or symmetry, fear of making mistakes, fear of sinful behavior, uncomfortable thoughts about sex or violence, good or bad numbers or colors, frightening mental images, fear of doing something uncontrolled, fear of negative effect of doing things the "wrong" way, and other.
5. Distress and impairment questions: After the symptoms had been queried, the interviewer asked about symptom-related distress and impairment. The distress and impairment questions were used at this point with the aim of ruling out subclinical symptoms as efficiently as possible. For example:
   a. "In the past month, how much does this problem bother or upset you?"
   b. "In the past month, does this problem make things harder for you, like make it hard to do well in school or at work, cause problems with your friends or family, or get in the way of doing fun or important things?"
   c. This was followed by a checklist of functional impairment domains: school, work or role functioning, social life, family, home responsibilities, leisure activities, legal problems, financial problems, problems of health or safety, or other functional impairment.
6. Clinical judgment ratings: Certain items did not require a specific question, but rather were to be rated by the interviewer based on all of the available information. In the case of OCD, these included Criteria C (the obsessive-compulsive symptoms are not attributable to the physiological effects of a substance or another medical condition) and D (the disturbance is not better explained by the symptoms of another mental disorder).
7. Information about differential diagnoses, associated features, and specifiers: To improve diagnostic accuracy among both experienced and novice interviewers, for each anxiety, mood, and obsessive-compulsive and related disorder, the DIAMOND-KID provided information about common differential diagnoses. For example, in the case of OCD, the DIAMOND-KID provides information (adapted from the *DSM-5*) about distinguishing OCD from anxiety disorders, depression, BDD, trichotillomania (TTM) and skin-picking/excoriation disorder (EXD), HD, eating disorders, tics and stereotyped movements, psychotic disorders, impulse control and SUDs, obsessive-compulsive personality disorder, and autism spectrum disorder. Associated features, also adapted from *DSM-5*, were listed, including typical dimensions of obsessions and compulsions; strong affective response when confronted with situations that trigger obsessions and compulsions; and avoidance of people, places, things, or activities that trigger

obsessions and compulsions. The possible specifiers for OCD were level of insight and the presence of tics.

Validity Measures. Participants also completed a battery of online self-report measures. The *Multidimensional Anxiety Scale for Children-2* (MASC-2; March, 2013) is a 50-item self-report measure of anxiety and related constructs. The MASC-2 subscales, which we used for convergent validity analyses, include Separation/Fears (α in this sample = 0.76), Generalized Anxiety (GAD) Index (α in this sample = 0.72), Obsessions/Compulsions (α in this sample = 0.87), Harm Avoidance (α in this sample = 0.60), Social Anxiety (Humiliation/Rejection and Performance Fears) (α in this sample = 0.84), and Physical Symptoms (Panic and Tense/Restless) (α in this sample = 0.82). The *Children's Depression Inventory-2* (CDI-2; Kovacs, 2011); α in this sample = 0.90) is a 28-item self-report measure of depression. The *Massachusetts General Hospital Hairpulling Scale* (MGH-HPS; Keuthen et al., 1995; α in this sample = 0.94) is a 7-item self-report measure of hairpulling behaviors. The *Skin Picking Scale* (SPS; Keuthen et al., 2001; α in this sample = 0.92) is a six-item self-report measure of skin-picking behaviors. The *ADHD Rating Scale-IV* (ARS-IV; Barkley & Murphy, 1998; α in this sample = 0.95) is an 18-item self-report measure of ADHD symptoms, for which we used the total score.[1]

Procedure

Participants (N = 311 children aged 10–17) were initially assessed using the DIAMOND-KID during an intake for clinical treatment. Interviews were administered by graduate students in clinical psychology, predoctoral psychology interns, postdoctoral fellows, or licensed psychologists. All interviewers received video-based training on the adult DIAMOND and were required to demonstrate good interrater reliability ($\kappa \geq 0.80$) with the trainer prior to clinical use over several video vignettes. However, no criterion was set in terms of the number of interviews given prior to data collection.

The initial interview was administered as part of standard clinical intake procedures. Within 1 week prior to the clinical interview, participants completed the MASC-2 and CDI-2 online and provided basic demographic information using Research Electronic Data Capture (REDCap) electronic data capture tools (Harris et al., 2009). The initial interview could be conducted with the child alone, or with the child and a parent/guardian, at the clinician's discretion (most were with the child alone). Subsequent interviews were with the child alone.

After the initial clinical interview, participants were invited to participate in the second phase of the study. Those agreeing to participate provided assent, and parents or guardians signed informed consent for research. They were scheduled to be interviewed again by telephone within approximately 2 weeks (M = 9.58 days, SD = 8.68) by a second interviewer (interrater reliability), and a third time by telephone approximately 1 week later (M = 8.09 days, SD = 4.32) by that same interviewer (test–retest reliability). Participants could opt to complete both the interrater and test–retest interviews or only one of these (the large majority completed both interviews; those who completed only one interview usually did so because of the inability to schedule the interview within the necessary time period). Prior to the reliability interviews, assenting participants completed the remaining self-report measures on REDCap. Participants were reimbursed in the form of $20 gift cards for each of the reliability interviews.

Interrater reliability was determined by calculating κ coefficients for each diagnosis (present or absent) between the initial assessment (rater 1) and first follow-up (rater 2). Test–retest reliability was determined by calculating κ coefficients for each diagnosis (present or absent) between the first follow-up (Rater 2) and the second follow-up (Rater 2). Following recommendations for the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012), κ coefficients of 0.80 and above are considered "excellent"; from 0.60 to 0.79 "very good"; from 0.40 to 0.59 "good"; from 0.20 to 0.39 "questionable"; and below 0.20 "unacceptable" (though see Vanheule et al. (2014) for a critique). We further examined the extent to which each κ coefficient could be considered statistically precise. When the standard error of κ is high (resulting in a wide confidence interval), even if the value of κ is high, the true κ cannot be estimated with precision. We therefore used the guidelines from the *DSM-5* field trials[1] to define a statistically precise estimate of κ as those with a standard error of $\leq$0.1 and a 95% confidence interval (CI) $\leq$0.5. For validity estimates, we used between-groups t tests and between-group effect size estimates (Cohen's d), using the presence and absence of each anxiety, mood, and OCD and related disorder as the independent variables, and scores on the corresponding self-report measures as the dependent variables.

Results

Interrater Reliability

Table 3 shows interrater reliability coefficients for the DIAMOND-KID diagnoses (presence vs. absence). κ coefficients ranged from questionable to very good, according to interpretive cutoffs used in the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012). The κ coefficients in the questionable range were GAD and MDD as well as any anxiety disorder (the latter was likely due to the questionable reliability of the GAD diagnosis). Only the κ

Table 3. Interrater Reliability for DIAMOND-KID Diagnoses.

	Diagnosis	K	95% CI	*t*	Interpretation
Anxiety disorders	SoP	0.60	[0.39, 0.80]	4.82**	Very good
	GAD	0.34	[0.10, 0.57]	2.74*	Questionable
	SpP	0.52	[0.29, 0.76]	4.24**	Good
	Any	0.26	[0.02, 0.51]	2.18*	Questionable
Depressive disorders	PDD	0.40	[0.14, 0.67]	3.53**	Good
	MDD	0.36	[0.11, 0.61]	2.89*	Questionable
	Any	0.56	[0.35, 0.76]	4.50**	Good
Obsessive-compulsive and related disorders	OCD	0.64	[0.45, 0.84][a]	5.20**	Very good
	TTM or EXD	0.52	[0.26, 0.78]	4.19**	Good
	Any	0.60	[0.41, 0.79][a]	4.86**	Very good
Other disorders	ADHD	0.69	[0.47, 0.90]	5.51**	Very good

Note. DIAMOND-KID = Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders: Child and Adolescent Version; SoP = social anxiety disorder (social phobia); GAD = generalized anxiety disorder; SpP = specific phobia; PDD = persistent depressive disorder (dysthymia); MDD = major depressive disorder; OCD = obsessive-compulsive disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; ADHD = attention-deficit/hyperactivity disorder; CI = confidence interval.
[a]Standard error ≤ 0.1 and 95% CI ≤ 0.5, indicating an acceptable estimate of κ.
*$p < .05$. **$p < .001$.

coefficient for OCD had both a standard error ≤ 0.1 and a 95% CI ≤ 0.5, suggesting a precise estimate using the definition from the field trial (Clarke et al., 2013).

Test–Retest Reliability

Table 4 shows test–retest reliability coefficients for the DIAMOND-KID diagnoses (presence vs. absence). For all diagnoses, κ coefficients ranged from very good to excellent, according to cutoffs used in the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012). For *DSM-5* categories, test–retest reliability for any anxiety disorder was very good. Test–retest reliability for any depressive or OCD and related disorder was excellent. Most κ coefficients had both a standard error ≤ 0.1 and a 95% CI ≤ 0.5, suggesting a precise estimate.

Convergent Validity

Table 5 shows self-report measure scores for participants with and without the corresponding diagnosis (based on the initial interview). Of note, the sample size was smaller for trichotillomania, skin picking disorder, and ADHD because the corresponding self-report measures were not given as part of routine clinical practice and therefore were only completed by children agreeing to participate in the interrater and test–retest reliability interviews. Between-group *t*-tests were significant for all diagnoses. Effect size estimates (Cohen's *d*) ranged from small (for participants with vs. without GAD on the GAD subscale of the MASC-2) to large (separation anxiety disorder [SAD], SoP, panic disorder, persistent depressive disorder, MDD, OCD, TTM, EXD, and ADHD).

Discussion

The aim of this study was to provide a preliminary analysis of the psychometric properties of a new semi-structured clinical interview for *DSM-5* anxiety, mood, OCD and related disorders for children and adolescents. Overall, the DIAMOND showed promising psychometric properties, with very good to excellent test–retest reliability using cutoff criteria from the *DSM-5* field trials (Clarke et al., 2013; Kraemer et al., 2012).

The present results may help clarify the reliability and validity of *DSM-5* diagnoses in children and adolescents. Of course, the psychometric properties of the *DSM*-5 cannot be disentangled from those of the interviews used to measure it. The *DSM-5* field trials (Clarke et al., 2013) employed a checklist-style interview which, to our knowledge, has not been published. Test–retest reliability is the only reliability estimate published from the *DSM-5* field trials, and those coefficients ranged from unacceptable to questionable for child anxiety and mood disorders (Regier et al., 2013). Within the published diagnostic interviews, the most frequently examined has been the K-SADS (Kaufman et al., 1997). In a small study in which parents and children were interviewed separately, the *DSM-IV* K-SADS showed acceptable inter-rater and test–retest reliability for a narrow range of conditions (Kaufman et al., 1997). Little is known about the psychometric properties of the *DSM-5* version of the K-SADS, and the existing data come primarily from smaller studies of translated versions of the measures which show variable reliability and validity for a narrow range of conditions (de la Pena et al., 2018; Nishiyama et al., 2020; Thornoretharson et al., 2020). The present results add to those findings by suggesting that

Table 4. Test–Retest Reliability for DIAMOND-KID Diagnoses.

	Diagnosis	κ	95% CI	*t*	Interpretation
Anxiety disorders	SoP	0.75	[0.58, 1.92][a]	5.71**	Very good
	GAD	0.72	[0.54, 0.91][a]	5.63**	Very good
	SpP	0.83	[0.64, 1.02][a]	6.30**	Excellent
	Any	0.78	[0.60, 0.96][a]	5.95**	Very good
Depressive disorders	PDD	0.76	[0.56, 0.96]	5.85**	Very good
	MDD	0.70	[0.50, 0.89][a]	5.36**	Very good
	Any	0.83	[0.68, 0.97][a]	6.30**	Excellent
Obsessive-compulsive and related disorders	OCD	0.81	[0.66, 0.97][a]	6.22**	Excellent
	TTM or EXD	0.85	[0.68, 1.02][a]	6.46**	Excellent
	Any	0.86	[0.73, 0.99][a]	6.56**	Excellent
Other disorders	ADHD	0.91	[0.79, 1.03][a]	6.87**	Excellent

Note. DIAMOND-KID = Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders: Child and Adolescent Version; SoP = social anxiety disorder (social phobia); PD = panic disorder; GAD = generalized anxiety disorder; SpP = specific phobia; PDD = persistent depressive disorder (dysthymia); MDD = major depressive disorder; OCD = obsessive-compulsive disorder; TTM = trichotillomania; EXD = excoriation (skin-picking) disorder; ADHD = attention-deficit/hyperactivity disorder; CI = confidence interval.
[a]Standard error ≤ 0.1% and 95% CI ≤ 0.5, indicating an acceptable estimate of κ.
*$p < .05$. **$p < .001$.

Table 5. Comparisons on Self-Report Measures for Participants With or Without Specific DIAMOND-KID Diagnoses.

	Diagnosis	Measure	With diagnosis	Without diagnosis	*t*	*d*
Anxiety disorders	SAD	MASC-SEP	15.67 (5.36) ($n = 9$)	8.91 (5.35) ($n = 295$)	3.73**	1.26
	SoP	MASC-SOC	18.60 (5.36) ($n = 112$)	11.22 (6.17) ($n = 192$)	10.54**	1.25
	PD	MASC-PAN	12.30 (4.96) ($n = 23$)	7.43 (4.86) ($n = 281$)	4.62**	1.00
	AGO	MASC-PAN	10.53 (4.50) ($n = 17$)	7.63 (5.02) ($n = 287$)	2.32*	0.58
	GAD	MASC-GAD	16.68 (5.59) ($n = 93$)	15.02 (5.36) ($n = 211$)	2.46*	0.31
	Any	MASC	70.68 (22.06) ($n = 197$)	60.34 (26.73) ($n = 107$)	3.62**	0.43
Depressive disorders	PDD	CDI-2	24.75 (6.55) ($n = 39$)	15.71 (8.90) ($n = 261$)	6.10**	1.05
	MDD	CDI-2	22.45 (8.14) ($n = 76$)	15.00 (8.70) ($n = 224$)	6.56**	0.87
	Any	CDI-2	22.99 (7.75) ($n = 112$)	13.24 (7.90) ($n = 188$)	10.41**	1.24
OCD and related disorders	OCD	MASC-OCD	15.54 (7.47) ($n = 125$)	7.60 (5.65) ($n = 179$)	10.55**	1.23
	TTM	MGH-HPS	12.40 (3.78) ($n = 5$)	1.13 (2.98) ($n = 63$)	8.01**	3.72
	EXD	SPS	11.30 (5.68) ($n = 10$)	2.67 (3.66) ($n = 58$)	6.31**	2.16
Other disorders	ADHD	ARS-IV	31.58 (14.58) ($n = 15$)	13.55 (9.21) ($n = 53$)	5.82**	1.70

Note. DIAMOND-KID = Diagnostic Interview for Anxiety, Mood, and OCD and Related Neuropsychiatric Disorders: Child and Adolescent Version; SAD = separation anxiety disorder; MASC-SEP = Multidimensional Anxiety Scale for Children-Separation Anxiety/Phobias Scale; SoP = social anxiety disorder (social phobia); MASC-SOC = Multidimensional Anxiety Scale for Children-Social Anxiety Scale; PD = panic disorder; MASC-PAN = Multidimensional Anxiety Scale for Children-Panic Scale; AGO = agoraphobia; GAD = generalized anxiety disorder; MASC-GAD = Multidimensional Anxiety Scale for Children-GAD Scale; PDD = persistent depressive disorder (dysthymia); CDI-2 = Children's Depression Inventory; MDD = major depressive disorder; OCD = obsessive-compulsive disorder; MASC-OCD = Multidimensional Anxiety Scale for Children-OCD Scale; TTM = trichotillomania; MGH-HPS = Massachusetts General Hospital Hair Pulling Scale; EXD = excoriation (skin-picking) disorder; SPS = Skin-Picking Scale; SUD = substance use disorder; ADHD = attention-deficit/hyperactivity disorder; ARS-IV = ADHD Rating Scale-IV; BDD = body dysmorphic disorder; HD = hoarding disorder; CI = confidence interval.
*$p < .05$. **$p < .001$.

DSM-5 diagnoses are reasonably stable over a 1-week time frame in children and adolescents.

It is noted, however, that interrater reliability statistics for GAD and MDD were in the questionable range. Interrater reliability was not assessed during the *DSM-5* field trials, and therefore there is no adequate basis for comparison. We note that one prior study of a translation of the *DSM-5* K-SADS showed acceptable interrater reliability for the GAD and MDD diagnoses (Thornoretharson et al., 2020); however, in that study reliability was assessed by having a second rater re-rate a recording of the initial interview, rather than by conducting a second interview as was done in the present study. The adult DIAMOND study (Tolin et al., 2018) did obtain inter-rater reliability coefficients; unlike the DIAMOND-KID, the adult DIAMOND GAD and MDD coefficients were in the very good range, suggesting that the lower reliability of those diagnoses may be specific to pediatric populations. As noted in the Introduction, a measure can only be as reliable and valid as the construct it purports to measure, and there is a question about whether the *DSM-5* diagnoses themselves are reliable and valid in children and adolescents. Assuming (for the moment) that the GAD and MDD diagnoses are inherently reliable and valid, we consider limitations of the DIAMOND administration that could account for the apparent discrepancy between the adult DIAMOND and the DIAMOND-KID. One possible explanation is that both GAD and MDD diagnoses are dependent on a specific time frame (6 months for GAD, 2 weeks for MDD), and younger interviewees may have difficulty reporting reliably on these time frames. We note again that although the first interview could have included both children and parents/guardians, the second and third interviews involved only the child, which is a limitation of the study as the DIAMOND-KID is meant to be a flexible tool that can take collateral reports into consideration. This underscores the potential importance of obtaining collateral reports (e.g., from parents/guardians) when possible and for further research to examine the additive benefit of such reports. Prior research has suggested low parent–child agreement on diagnostic measures (Orchard et al., 2019; Rapee et al., 1994), making multi-informant data collection a critical issue for further study (De Los Reyes et al., 2015). Future research should investigate the additive benefit of including parents or guardians in the interview process and determine whether that benefit is age- or diagnosis-dependent. Furthermore, unlike some other interviews which can be administered separately to children and to parents/guardians, the DIAMOND-KID is structured to allow both in the room at the same time, and it would be informative to compare the psychometric properties of separate vs. joint interviews.

Convergent validity of the DIAMOND-KID diagnoses was verified by higher scores on corresponding self-report measures for participants with specific anxiety, mood, and OCD and related diagnoses. Thus, the DIAMOND-KID diagnoses appear to measure their intended constructs.

One important limitation of the present study is the underrepresentation of certain target disorders including SM, bipolar disorders, DMDD, PMDD, BDD, and HD, due to very low prevalence in the clinical sites used for the study. Another limitation is the fact that although we assessed the reliability of diagnoses (present vs. absent), we did not examine the reliability of the designation of a diagnosis as principal. In the current data, multiple diagnoses received the same severity score and therefore principal diagnoses cannot be inferred from those scores. Future research should include primary diagnosis as a variable of interest.

The use of telephone interviews is a further limitation of the present study. Some DIAMOND-KID diagnoses (e.g., selective mutism or mood disorders) can incorporate behavioral observations by the clinician, which is likely hampered in a telephone interview. Future research should use in-person interviews, and clarify the reliability of a telephone vs. in-person format.

The reliance on child self-report for both symptom information (DIAMOND-KID) and the convergent validity measures is also a limitation. Although the reliability of reports by children aged 8 and older is good on child-focused health questionnaires (Riley, 2004), and child mental health measures show good reliability (e.g., Reynolds & Graves, 1989; Ridge et al., 2009), the addition of parent report of child symptoms would increase confidence in the present findings. We note as well that a strong test of validity would be provided by comparing DIAMOND-KID diagnoses to those obtained by other structured interviews; however, as described in the Introduction and depicted in Table 1, there is no single, suitable "gold standard" interview with which to compare the DIAMOND-KID, especially concerning the OCD and related disorders.

We note that children aged 6 to 9 were underrepresented and not analyzed here due to the low number of children in that age range. Thus, the DIAMOND-KID is reliable for children older than the age of 10, while the reliability in younger children is unclear. Additional research is needed to examine the reliability and validity of the DIAMOND-KID in younger children, likely with input from parents and guardians.

While we consider our use of different interviews for inter-rater and test–retest reliability, rather than re-rating audio recordings as has been done previously (Thornoretharson et al., 2020), as a strength, we recognize that both approaches have their limitations. Re-rating audio recordings may artificially inflate reliability estimates, and this would be particularly true for an instrument such as the DIAMOND-KID which "skips out" when diagnostic criteria are not met. In such a case, the re-rater could simply listen for which diagnoses "skipped out" and rate accordingly.

However, it could also be argued that conducting entirely new interviews, as was done here and as was done with the adult DIAMOND (Tolin et al., 2018), confounds the reliability of administration with different responses and different informants, potentially diminishing reliability estimates. The present findings, therefore, are likely to represent a conservative estimate of interrater and test–retest reliability.

In summary, the DIAMOND-KID appears to be a promising measure of anxiety, mood, and obsessive-compulsive and related disorders in children and adolescents, although we suspect that younger children's interviews will be more reliable if a parent or guardian can be involved. DIAMOND-KID diagnoses comport well with scores on self-report measures, and test–retest reliability is very good to excellent. These findings further suggest that the *DSM-5* diagnoses targeted by this measure are themselves reasonably reliable and valid.

Acknowledgment

We thank Dr. Moselle Campbell, Terence Ching, Dr. Carolyn Davies, Dr. Gretchen Diefenbach, Anders Hogstrom, Dr. Eric Lee, Dr. Hannah Levy, Monica Ly, Dr. Emily O'Bryan, and Kiera Valentine for conducting DIAMOND-KID interviews.

Authors' Note

Copies of the DIAMOND-KID may be obtained at no cost from the first author at david.tolin@hhchealth.org or at https://giving.harthosp.org/tolin-diamond-training-video.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iDs

David F. Tolin https://orcid.org/0000-0003-0123-889X

Sabrina Darrow https://orcid.org/0000-0002-5639-0285

Marisa D. Serchuk https://orcid.org/0000-0002-3787-5957

Note

1. We also administered the *Mood Disorder Questionnaire* (MDQ; Hirschfeld et al., 2000), a screening measure for bipolar disorder; the *Body Dysmorphic Disorder Questionnaire* (BDDQ; Phillips, 1996) a self-report measure of body dissatisfaction and preoccupation; the *PTSD Diagnostic Scale for Children* (PDS-C; Foa et al., 2001), a self-report measure of PTSD symptoms; the *Children's Eating Attitudes Test* (ChEAT; Maloney et al., 1988) a self-report measure of disordered eating-related beliefs and behaviors, and the *CRAFFT Screening Tool* (Knight et al., 1999), a self-report measure of substance abuse. Because of the relatively low number of participants with diagnoses assessed by these measures, they will not be discussed further in the present article.

References

Aebi, M., Kuhn, C., Metzke, C. W., Stringaris, A., Goodman, R., & Steinhausen, H. C. (2012). The use of the development and well-being assessment (DAWBA) in clinical practice: A randomized trial. *European Child and Adolescent Psychiatry*, *21*(10), 559–567. https://doi.org/10.1007/s00787-012-0293-6

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th text rev., ed.).

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.).

Angold, A., Prendergast, M., Cox, A., Harrington, R., Simonoff, E., & Rutter, M. (1995). The Child and Adolescent Psychiatric Assessment (CAPA). *Psychological Medicine*, *25*(4), 739–753. https://doi.org/10.1017/s003329170003498x

Arifin, W. N. (2022). *Sample size calculator*. http://wnarifin.github.io

Barkley, R. A., & Murphy, K. R. (1998). *Attention-deficit hyperactivity disorder: A clinical workbook* (2nd ed.). Guilford Press.

Brown, T. A., & Barlow, D. H. (2009). A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment*, *21*(3), 256–271. https://doi.org/10.1037/a0016608

Clarke, D. E., Narrow, W. E., Regier, D. A., Kuramoto, S. J., Kupfer, D. J., Kuhl, E. A., Greiner, L., & Kraemer, H. C. (2013). DSM-5 field trials in the United States and Canada, Part I: Study design, sampling strategy, implementation, and analytic approaches. *American Journal of Psychiatry*, *170*(1), 43–58. https://doi.org/10.1176/appi.ajp.2012.12070998

Cuthbert, B. N. (2014). The RDoC framework: Facilitating transition from ICD/DSM to dimensional approaches that integrate neuroscience and psychopathology. *World Psychiatry*, *13*(1), 28–35. https://doi.org/10.1002/wps.20087

de la Pena, F. R., Villavicencio, L. R., Palacio, J. D., Felix, F. J., Larraguibel, M., Viola, L., Ortiz, S., Rosetti, M., Abadi, A., Montiel, C., Mayer, P. A., Fernandez, S., Jaimes, A., Feria, M., Sosa, L., Rodriguez, A., Zavaleta, P., Uribe, D., Galicia, F., & Ulloa, R. E. (2018). Validity and reliability of the Kiddie Schedule for Affective Disorders and Schizophrenia present and lifetime version DSM-5 (K-SADS-PL-5) Spanish version. *BMC Psychiatry*, *18*(1), Article 193. https://doi.org/10.1186/s12888-018-1773-0

De Los Reyes, A., Augenstein, T. M., Wang, M., Thomas, S. A., Drabick, D. A. G., Burgers, D. E., & Rabinowitz, J. (2015). The validity of the multi-informant approach to assessing child and adolescent mental health. *Psychological Bulletin*, *141*(4), 858–900. https://doi.org/10.1037/a0038498

Dyl, J., Kittler, J., Phillips, K. A., & Hunt, J. I. (2006). Body dysmorphic disorder and other clinically significant body image concerns in adolescent psychiatric inpatients: Prevalence and clinical characteristics. *Child Psychiatry and Human Development*, *36*(4), 369–382. https://doi.org/10.1007/s10578-006-0008-7

First, M. B. (2005). Clinical utility: A prerequisite for the adoption of a dimensional approach in DSM. *Journal of Abnormal Psychology*, *114*(4), 560–564. https://doi.org/10.1037/0021-843X.114.4.560

Foa, E. B., Johnson, K. M., Feeny, N. C., & Treadwell, K. R. (2001). The child PTSD Symptom Scale: A preliminary examination of its psychometric properties. *Journal of Clinical Child Psychology*, *30*(3), 376–384. https://doi.org/10.1207/S15374424JCCP3003_9

Harris, P. A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., & Conde, J. G. (2009). Research electronic data capture (REDCap)—A metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*, *42*(2), 377–381. https://doi.org/10.1016/j.jbi.2008.08.010

Hirschfeld, R. M., Williams, J. B., Spitzer, R. L., Calabrese, J. R., Flynn, L., Keck, P. E., Jr., Lewis, L., McElroy, S. L., Post, R. M., Rapport, D. J., Russell, J. M., Sachs, G. S., & Zajecka, J. (2000). Development and validation of a screening instrument for bipolar spectrum disorder: The Mood Disorder Questionnaire. *American Journal of Psychiatry*, 157(11), 1873–1875. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11058490

Hojgaard, D., Skarphedinsson, G., Ivarsson, T., Weidle, B., Nissen, J. B., Hybel, K. A., Torp, N. C., Melin, K., & Thomsen, P. H. (2019). Hoarding in children and adolescents with obsessive-compulsive disorder: Prevalence, clinical correlates, and cognitive behavioral therapy outcome. *European Child and Adolescent Psychiatry*, *28*(8), 1097–1106. https://doi.org/10.1007/s00787-019-01276-x

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., Williamson, D., & Ryan, N. (1997). Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*(7), 980–988. https://doi.org/10.1097/00004583-199707000-00021

Kendell, R., & Jablensky, A. (2003). Distinguishing between the validity and utility of psychiatric diagnoses. *American Journal of Psychiatry*, *160*(1), 4–12. https://doi.org/10.1176/appi.ajp.160.1.4

Keuthen, N. J., O'Sullivan, R. L., Ricciardi, J. N., Shera, D., Savage, C. R., Borgmann, A. S., Jenike, M. A., & Baer, L. (1995). The Massachusetts General Hospital (MGH) Hairpulling Scale: 1. Development and factor analyses. *Psychotherapy and Psychosomatics*, 64(3–4), 141–145. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=8657844

Keuthen, N. J., Wilhelm, S., Deckersbach, T., Engelhard, I. M., Forker, A. E., Baer, L., & Jenike, M. A. (2001). The Skin Picking Scale: Scale construction and psychometric analyses. *Journal of Psychosomatic Research*, 50(6), 337–341. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11438115

Knight, J. R., Shrier, L. A., Bravender, T. D., Farrell, M., Vander Bilt, J., & Shaffer, H. J. (1999). A new brief screen for adolescent substance abuse. *Archives of Pediatrics and Adolescent Medicine*, *153*(6), 591–596. https://doi.org/10.1001/archpedi.153.6.591

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., & Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kovacs, M. (2011). *Children's Depression Inventory 2nd edition: Technical manual*. Multi-Health Systems.

Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). DSM-5: How reliable is reliable enough? *American Journal of Psychiatry*, *169*(1), 13–15.

Krueger, R. F., & Bezdjian, S. (2009). Enhancing research and treatment of mental disorders with dimensional concepts: Toward DSM-V and ICD-11. *World Psychiatry*, 8(1), 3–6. http://www.ncbi.nlm.nih.gov/pubmed/19293948

Krueger, R. F., Markon, K. E., Patrick, C. J., & Iacono, W. G. (2005). Externalizing psychopathology in adulthood: A dimensional-spectrum conceptualization and its implications for DSM-V. *Journal of Abnormal Psychology*, *114*(4), 537–550. https://doi.org/10.1037/0021-843X.114.4.537

Leffler, J. M., Riebel, J., & Hughes, H. M. (2015). A review of child and adolescent diagnostic interviews for clinical practitioners. *Assessment*, *22*(6), 690–703. https://doi.org/10.1177/1073191114561253

Maloney, M. J., McGuire, J. B., & Daniels, S. R. (1988). Reliability testing of a children's version of the Eating Attitude Test. *Journal of the American Academy of Child and Adolescent Psychiatry*, *27*(5), 541–543. https://doi.org/10.1097/00004583-198809000-00004

March, J. S. (2013). *Multidimensional Anxiety Scale for Children 2nd edition: Technical manual*. Multi-Health Systems.

Moreno-Amador, B., Cervin, M., Falcó, R., Marzo, J. C., & Piqueras, J. A. (2022). Body-dysmorphic, hoarding, hair-pulling, and skin-picking symptoms in a large sample of adolescents. *Current Psychology*, 1–12.

Nishiyama, T., Sumi, S., Watanabe, H., Suzuki, F., Kuru, Y., Shiino, T., Kimura, T., Wang, C., Lin, Y., Ichiyanagi, M., & Hirai, K. (2020). The Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version (K-SADS-PL) for DSM-5: A validation for neurodevelopmental disorders in Japanese outpatients. *Comprehensive Psychiatry*, *96*, Article 152148. https://doi.org/10.1016/j.comppsych.2019.152148

Orchard, F., Pass, L., Cocks, L., Chessell, C., & Reynolds, S. (2019). Examining parent and child agreement in the diagnosis of adolescent depression. *Child and Adolescent Mental Health*, *24*(4), 338–344. https://doi.org/10.1111/camh.12348

Phillips, K. A. (1996). *The broken mirror: Understanding and treating body dysmorphic disorder*. Oxford University Press.

Rapee, R. M., Barrett, P. M., Dadds, M. R., & Evans, L. (1994). Reliability of the DSM-III-R childhood anxiety disorders using structured interview: Interrater and parent-child agreement. *Journal of the American Academy of Child and Adolescent Psychiatry*, 33(7), 984–992. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=7961354

Regier, D. A., Narrow, W. E., Clarke, D. E., Kraemer, H. C., Kuramoto, S. J., Kuhl, E. A., & Kupfer, D. J. (2013). DSM-5 field trials in the United States and Canada, Part II: Test-retest reliability of selected categorical diagnoses. *American Journal of Psychiatry*, *170*(1), 59–70. https://doi.org/10.1176/appi.ajp.2012.12070999

Reynolds, W. M., & Graves, A. (1989). Reliability of children's reports of depressive symptomatology. *Journal of Abnormal Child Psychology*, *17*(6), 647–655. https://doi.org/10.1007/BF00917728

Ridge, N. W., Warren, J. S., Burlingame, G. M., Wells, M. G., & Tumblin, K. M. (2009). Reliability and validity of the youth outcome questionnaire self-report. *Journal of Clinical Psychology*, *65*(10), 1115–1126. https://doi.org/10.1002/jclp.20620

Riley, A. W. (2004). Evidence that school-age children can self-report on their health. *Ambulatory Pediatrics*, *4*(4), 371–376.

Sheehan, D. V., Sheehan, K. H., Shylte, R. D., Janavs, J., Bannon, Y., Rogers, J. E., Milo, K. M., Stock, S. L., & Wilkinson, B. (2010). Reliability and validity of the Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID). *Journal of Clinical Psychiatry*, *71*(3), 313–326.

Stroud, J. D. (1983). Hair loss in children. *Pediatric Clinics of North America*, *30*(4), 641–657. https://doi.org/10.1016/s0031-3955(16)34431-5

Thornoretharson, O., Ævarsson, F. M., Helgadottir, S., Lauth, B., Wessman, I., Sigurjonsdottir, S. A., Smarason, O., Harethardottir, H. H., & Skarphedinsson, G. (2020). Icelandic translation and reliability data on the DSM-5 version of the Schedule for Affective Disorders and Schizophrenia for School-Aged Children—Present and Lifetime Version (K-SADS-PL). *Nordic Journal of Psychiatry*, *74*(6), 423–428. https://doi.org/10.1080/08039488.2020.1733660

Tolin, D. F., Gilliam, C., Wootton, B. M., Bowe, W., Bragdon, L. B., Davis, E., Hannan, S. E., Steinman, S. A., Worden, B., & Hallion, L. S. (2018). Psychometric properties of a structured diagnostic interview for DSM-5 anxiety, mood, and obsessive-compulsive and related disorders. *Assessment*, *25*(1), 3–13. https://doi.org/10.1177/1073191116638410

Tyrer, P. (2014). Time to choose: DSM-5, ICD-11 or both? *Archives of Psychiatry and Psychotherapy*, *3*, 5–8.

Vanheule, S., Desmet, M., Meganck, R., Inslegers, R., Willemsen, J., De Schryver, M., & Devisch, I. (2014). Reliability in psychiatric diagnosis with the DSM: Old wine in new barrels. *Psychotherapy and Psychosomatics*, *83*(5), 313–314. https://doi.org/10.1159/000358809

Young, M. E., Bell, Z. E., & Fristad, M. A. (2016). Validation of a brief structured interview: The Children's Interview for Psychiatric Syndromes (ChIPS). *Journal of Clinical Psychology in Medical Settings*, *23*(4), 327–340. https://doi.org/10.1007/s10880-016-9474-7