J Clin Psychol Med Settings (2016) 23:327–340
DOI 10.1007/s10880-016-9474-7

ORIGINAL PAPER

Validation of a Brief Structured Interview: The Children's Interview for Psychiatric Syndromes (ChIPS)

Matthew E. Young[1] · Ziv E. Bell[2] · Mary A. Fristad[3]

Published online: 19 October 2016

Abstract Evidence-based assessment is important in the treatment of childhood psychopathology. While researchers and clinicians frequently use structured diagnostic interviews to ensure reliability, the most commonly used instrument, the Schedule for Affective Disorders and Schizophrenia for School Aged Children (K-SADS) is too long for most clinical applications. The Children's Interview for Psychiatric Syndromes (ChIPS/P-ChIPS) is a highly-structured brief diagnostic interview. The present study compared K-SADS and ChIPS/P-ChIPS diagnoses in an outpatient clinical sample of 50 parent–child pairs aged 7–14. Agreement between most diagnoses was moderate to high between the instruments and with consensus clinical diagnoses. ChIPS was significantly briefer to administer than the K-SADS. Interviewer experience level and participant demographics did not appear to affect agreement. Results provide further evidence for the validity of the ChIPS and support its use in clinical and research settings.

Keywords Structured diagnostic interview · Assessment · Validity · Psychometrics

Validation of a Brief Structured Interview: Ready for Prime Time Clinical Practice

In the past decade, evidence-based treatments (EBTs) for children and adolescents have generated increased attention in both research and clinical contexts (e.g., Chorpita et al., 2011). Over this same period, comparatively less research has been devoted to the development of evidence-based assessment (EBA) methods for children and adolescents. This oversight requires attention because the validity of EBTs depends on carefully conducted and empirically-supported assessments, and many commonly-used assessment practices are not supported by evidence (Jensen-Doss, 2011). EBA can occur at numerous time points and can also inform each stage of EBT. For example, an initial diagnostic assessment can determine whether an EBT is appropriate for a particular child or adolescent, and if so, which one (Youngstrom, 2013). During treatment, EBA can be used to monitor progress, make adjustments as needed, and quantify treatment effects.

Many diagnostic interviews have been developed for use with children and adolescents. Available measures differ with regard to their format; informants utilized (e.g., parent, child, both parent and child, siblings, teachers); and the diagnostic system utilized (e.g., DSM-5; American Psychiatric Association, 2013). Figure 1 summarizes the relevant characteristics of these approaches to collecting clinical diagnostic information. In summary, semi-structured interviews tend to be longer and have greater training requirements than their highly-structured counterparts.

✉ Mary A. Fristad
mary.fristad@osumc.edu

Matthew E. Young
Myoung7@uchicago.edu

Ziv E. Bell
Bell.1344@buckeyemail.osu.edu

[1] Department of Psychiatry and Behavioral Neuroscience, The University of Chicago Medicine, Chicago, IL, USA

[2] Department of Psychology, The Ohio State University, Columbus, OH, USA

[3] Department of Psychiatry and Behavioral Health, The Ohio State Wexner Medical Center, 1670 Upham Drive Suite 460G, Columbus, OH 43210-1250, USA

A variety of semi-structured and structured diagnostic interviews are available for use with children and adolescents and their parents, including the Child and Adolescent Psychiatric Assessment (CAPA; Angold & Costello, 2000); the Diagnostic Interview Schedule for Children (DISC; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000); and the Diagnostic Interview for Children and Adolescents (DICA; Reich, 2000). These interviews serve a multitude of purposes (e.g. Rolon-Arroyo, Arnold, Harvey, & Marshall, 2016), but most were developed for research use in child and adolescent psychiatry and are not convenient for clinical application due to their length (Angold et al., 2012). The Mini International Neuropsychiatric Interview for Children and Adolescents (MINI-KID; Sheehan et al., 2010) is a brief diagnostic interview designed to address many of the barriers to use in clinical settings, however the MINI-KID assesses present diagnoses and only assesses lifetime symptoms of major depression, suicidality, and psychotic disorders. These measures have yet to be revised and validated for DSM-5. Standardization samples have been relatively small or confined to a single population (i.e., community or clinical, rather than both), and most interviews' diagnoses have not been compared to an external criterion (e.g., chart diagnoses). Although diagnostic interviews have the potential to standardize and streamline the diagnostic process for children and adolescents, existing empirical evidence does not strongly support the use of one interview over others (for a review, see Leffler, Riebel, & Hughes, 2015).

The most frequently used diagnostic interview is the Schedule for Affective Disorders and Schizophrenia for School Aged Children (K-SADS; Ambrosini, 2000; Puig-Antich & Chambers, 1978; Puig-Antich & Ryan, 1986). The K-SADS is commonly used in studies of childhood mood disorders and incorporates both child and parent reports. Each symptom also has a "summary score," assigned by the interviewer after considering child and parent reports and all other available sources of clinical information, such as behavioral observations, self-report questionnaires, and teacher ratings.

Available reliability and validity estimates for the most recent K-SADS versions range from acceptable to excellent, but are based on relatively small samples (Birmaher et al., 2006; Goldstein et al., 2005; Kim, Miklowitz, Biuckians, & Mullen, 2007; Lewinsohn, Shankman, Gau, & Klein, 2004). In general, inter-rater reliability is excellent across diagnoses and test–retest reliability is excellent among mood disorders, current anxiety disorders, oppositional defiant disorder (ODD), and lifetime conduct disorder (Kaufman et al., 1997). Test–retest reliability is moderate for lifetime ADHD, posttraumatic stress disorder (PTSD), and lifetime anxiety disorders. The K-SADS demonstrates good convergent and divergent validity with behavioral checklists (Birmaher et al., 2009). The K-SADS has been adapted for use in other languages and cultures (Akdemir & Gokler, 2008; Ghanizadeh, Mohammadi, & Yazdanshenas, 2006; Kim et al., 2004; Renou, Hergueta, Flament, Mouren-Simeoni, & Lecrubier, 2004; Schoeman, Carey, & Seedat, 2009), with some available data on cross-cultural K-SADS adaptations indicating high inter-rater reliability and validity (Brasil & Bordin, 2010; Kolaitis, Korpa, Kolvin, & Tsiantis, 2003; Lauth et al., 2010; Shanee, Apter, & Weizman, 1997).

A potential weakness of the K-SADS is that significant training and clinical judgment are required, compared to more structured interviews. Interviewers must have an understanding of child psychopathology and knowledge of DSM-IV diagnostic criteria. Training involves didactic and observation components, joint interviews to calibrate scoring, and periodic re-checks to maintain inter-rater reliability (Kaufman, et al., 1997). The interviewer must exercise clinical judgment throughout the K-SADS in ways

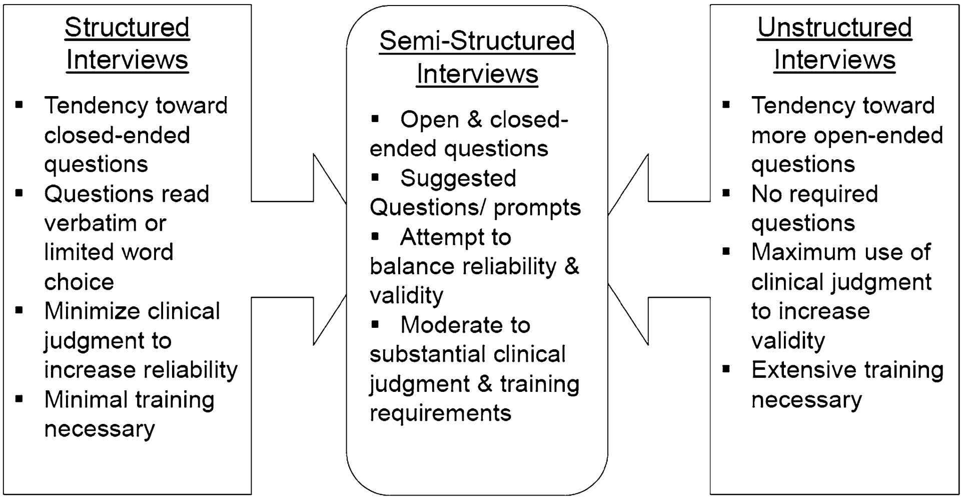
Structured Interviews
▪ Tendency toward closed-ended questions
▪ Questions read verbatim or limited word choice
▪ Minimize clinical judgment to increase reliability
▪ Minimal training necessary
Semi-Structured Interviews
▪ Open & closed-ended questions
▪ Suggested Questions/ prompts
▪ Attempt to balance reliability & validity
▪ Moderate to substantial clinical judgment & training requirements
Unstructured Interviews
▪ Tendency toward more open-ended questions
▪ No required questions
▪ Maximum use of clinical judgment to increase validity
▪ Extensive training necessary

Fig. 1 Comparison of clinical interviewing procedures

that can influence information collected. For example, the interviewer must choose which suggested probe questions to ask for each symptom, determine the order in which to ask them, and decide when to ask additional questions not included on the instrument.

The K-SADS interview is also time-consuming. It utilizes a screen interview followed by supplements for all disorders for which symptoms are endorsed. Therefore, interview length varies based on the number of symptoms reported and can last upwards of 180 min to administer to both parent and child (Ambrosini, 2000). Clinicians in most settings cannot devote this much time to a diagnostic assessment due to heavy caseloads and reimbursement practices associated with session length. Even though the K-SADS is widely used in research (Galanter, Hundt, Goyal, Le, & Fisher, 2012; Nottelmann et al., 2001; Van Meter, Moreira, & Youngstrom, 2011), it is rarely used in clinical settings (Youngstrom, Findling, Youngstrom, & Calabrese, 2005), likely due to practical limitations discussed above. The lack of consistent assessment practices across settings raises questions about whether research diagnoses differ from those assigned clinically. Empirical studies often find low agreement between clinician-assigned diagnoses from unstructured interviews and diagnoses from comprehensive, structured interviews (Jensen-Doss, Youngstrom, Youngstrom, Feeny, & Findling, 2014). For example, children with a K-SADS diagnosis of major depressive disorder (MDD) may experience more severe psychopathology than children diagnosed with MDD in routine clinical practice (Hamilton & Gilham, 1999). Given such differences, results of clinical research studies that utilize the K-SADS may not generalize well to clinical settings. To provide EBA and subsequent EBT, clinicians need an assessment tool with comparable psychometric properties and diagnostic utility, but fewer practical barriers for use in clinical practice.

The Children's Interview for Psychiatric Syndromes, child (ChIPS; Weller, Weller, Rooney, & Fristad, 1999a) and parent (P-ChIPS; Weller, Weller, Rooney, & Fristad, 1999b) versions, are highly-structured diagnostic interviews that assess 20 DSM-IV diagnoses. The ChIPS was developed following a review of the characteristics of existing structured interviews, and its questions were written with an emphasis on simple, age-appropriate language (Teare, Fristad, Weller, Weller, & Salmon, 1998a). A DSM-5 version has been developed and is currently undergoing psychometric evaluation (Fristad, personal communication, May 2, 2016). Like the K-SADS, the ChIPS includes ratings of clinical impairment and onset for each diagnosis. It uses a branching format, allowing the interviewer to "skip out" of a section as soon as it is clear that the child does not meet diagnostic criteria for that diagnosis. Symptoms are listed in the order of expected frequency, allowing "cardinal symptoms" to be inquired about first (Weller, Weller, Fristad, Rooney, & Schecter, 2000). Unlike more recent versions of the K-SADS, it does not have separate "past" sections for each symptom or diagnosis. Instead, the ChIPS inquires whether these problems "ever" occurred for the child, and the interviewer records onset and offset dates at the end of each section.

Administration is brief relative to the K-SADS. A ChIPS or P-ChIPS interview takes approximately 36 min per informant in an outpatient sample (Rooney, Fristad, Weller, & Weller, 1999). Because of the yes/no format of most items, the ChIPS requires limited clinical judgment, and non-clinical staff under clinical supervision can complete the interview. ChIPS was designed as a screening tool, so false positives are expected but few false negatives are anticipated. Given this, the ChIPS results are later interpreted by a clinician, who can determine the clinical importance of syndromes endorsed. Another potential advantage of the ChIPS compared to other instruments is that it requires relatively little storage space. The parent and child forms of the interview are published in a reusable spiral-bound format. The interviewer codes responses on an eight-page scoring form, and later records ChIPS results on a four-page summary report form for storage in a client or research participant's file or scanned into an electronic medical record. In contrast, the K-SADS screening interview alone can be over 100 pages long. Depending on the version used and the number of supplements administered, one child's K-SADS interview can be almost 200 pages long. In a busy clinic, storage requirements for the ChIPS can be considerably more cost-effective than for a longer, albeit potentially more exhaustive assessment instrument such as the K-SADS.

Reliability and validity of the ChIPS has been investigated in inpatient, outpatient, and community samples with both children and adolescents in a series of studies conducted by the measure's developers. The current version of the ChIPS demonstrated similar agreement with the DICA (Reich, 2000); the Youth Self Report (YSR; Ebesutani et al., 2010); and discharge diagnoses in a sample of 47 inpatient children and adolescents (Fristad, Cummins, et al., 1998). Sensitivity was 0.70 and specificity was 0.84 when compared to discharge diagnoses. Both the ChIPS ($M = 3.8$) and DICA ($M = 3.4$) assigned more diagnoses on average than clinicians ($M = 2.1$) in this sample. In a community sample of bereaved ($n = 18$) and control ($n = 22$) children and adolescents, the ChIPS demonstrated adequate agreement with clinician consensus on diagnoses and symptoms (Fristad, Glickman, et al., 1998). The proportion of children in this sample who received one or more diagnoses (17.5 %) was similar to the proportion expected in a non-clinical sample, and administration was brief (mean = 21 min). Across samples and studies, administration time was significantly shorter than the DICA (Teare, Fristad, Weller, Weller, & Salmon, 1998b).

Psychometric evidence supporting the ChIPS suggests that it is a reliable and valid instrument for assigning DSM-IV diagnoses in children and adolescents. Combining results of previous studies, Weller et al. (2000) noted both the ChIPS and P-ChIPS demonstrated a negative predictive value (i.e., the number of true negatives determined by clinicians divided by the number of negative cases generated by ChIPS) of .96. This reinforces its value as a screening measure, as diagnoses not endorsed on the ChIPS are unlikely to be present. In a small sample ($N = 12$), telephone administration of the P-ChIPS was comparable to face-to-face administration (Paing, Weller, Dixon, & Weller, 2010). ChIPS/P-ChIPS have superior estimates of sensitivity, child-clinician concordance, and parent-clinician concordance and equivalent estimates of specificity compared to published reports of similar structured interviews (Weller et al., 2000). The ChIPS has been used as a diagnostic measure in research focusing on topics such as childhood bipolar disorders (e.g., Fristad, 2006; Hunt et al., 2005), the genetics of reading disabilities (Luca et al., 2007), and the comorbidity between ADHD and depression (Quintana, Butterbaugh, Purnell, & Layman, 2007).

Like most highly structured interviews, the ChIPS requires relatively little clinical judgment for scoring, due to the yes/no format of many of its items. Therefore, interviewer training is relatively brief compared to semi-structured instruments. ChIPS training involves reading the administration manual, didactic training by an experienced interviewer, and joint interviews to calibrate scoring. Inter-rater reliability can be evaluated through periodic joint interviews or rating videotapes of interviews (Rooney et al., 1999). In summary, the ChIPS is relatively brief and inexpensive to administer and provides comprehensive coverage of diagnostic criteria for numerous DSM-IV disorders, making it an attractive option for clinical and research use (Leffler et al., 2015).

The ChIPS and K-SADS are both designed to assign diagnoses consistent with DSM-IV criteria, and can potentially be utilized in both clinical and research settings. Both measures, like all diagnostic interviews, are intended to increase the reliability of diagnosis relative to an unstructured clinical interview. However, several differences between the ChIPS and K-SADS highlight the importance of comparing the concurrent validity of these measures. The ChIPS limits the interviewer's clinical judgment with a relatively rigid format for ease and speed of administration and the intent of maximizing reliability. Comparatively, the K-SADS, though significantly more time consuming, offers interviewers the opportunity to collect more detained information, rate the severity of some symptom categories, and use their clinical judgment within the framework of the measure. The ChIPS requires fewer resources including time, interviewer training and skill level, and physical storage space. The K-SADS offers broader coverage and allows interviewers to ask additional information as they deem necessary, to reduce the need for later follow-up information and with the goal of maximizing clinical validity. Given these differences, a comparison of the ChIPS and K-SADS has potential implications for both research and clinical settings. If these two interviews demonstrate high rates of agreement on diagnosis, researchers could consider administering the ChIPS instead of the K-SADS. This would save time and money for research assessments, both in the length of administration and in the scope of interviewer training. It could also aid study recruitment and retention, as potential participants may be more agreeable to a shorter assessment. In clinical settings, the ChIPS could be inexpensively administered by less-experienced staff supervised by experienced clinicians. The clinician interpreting its results could be confident that the ChIPS diagnosis closely approximates diagnoses assigned in published literature.

To date, one published study has directly compared diagnostic agreement between these interviews. In this study, Swenson et al. (2007) examined concurrent validity between ChIPS diagnoses (child interview only) and K-SADS diagnoses (child interview only), as well as scores on a number of self-report symptom scales in patients age 12–18 on a psychiatric inpatient unit. This study found moderate agreement between ChIPS and K-SADS diagnoses (mean $\kappa = .43$; range: .18 to .66), and similar rates of agreement for both measures and the symptom rating scales (Swenson et al., 2007). However, this study excluded adolescents with psychosis, did not examine diagnoses of mania/hypomania, and did not include parent interviews. The study also did not compare diagnoses from either measure to an independent criterion, such as admission or chart diagnoses, so sensitivity and specificity calculations were precluded. Finally, the ChIPS was always administered at admission by a bachelor's level interviewer, whereas the K-SADS was administered an average of three days later, by a masters- or doctoral-level interviewer (Swenson et al., 2007). It is unclear whether interview order, interviewer education level, or the lack of a parental informant affected agreement between these interviews, and whether these results will generalize to outpatient samples or younger age groups.

Thus, in this study, the ChIPS and the K-SADS were compared in 50 youth with a range of psychopathology. Three hypotheses were evaluated. First, it was hypothesized that ChIPS diagnoses and K-SADS diagnoses would demonstrate a high rate of agreement (i.e. kappa $\geq .6$) for each diagnosis assessed by both instruments. Second, it was expected that rates of agreement with diagnosis assigned by a licensed clinician would not differ significantly from ChIPS and K-SADS diagnoses. Third, the mean duration of ChIPS interviews were expected to be significantly shorter than the K-SADS in an outpatient setting reflective of many community-based clinics and practices. Support of these hypotheses should empower clinicians and researchers to use an EBA in their work without incurring costs related to time, space, and training typical of many other EBAs.

Method

Participants

Participants were 50 children and adolescents (age 7–14), and one parental informant for each child (total $N = 100$ participants). Participants were recruited from families participating in the follow-up portion of the Longitudinal Assessment of Manic Symptoms (LAMS) study (RO1MH073801) at a Midwestern medical center. They had a wide range of DSM-IV disorders at study entry, including mood disorders, anxiety disorders, schizophrenia and other psychotic disorders, posttraumatic stress disorder, adjustment disorders, behavioral disorders (including ADHD), eating disorders, and elimination disorders; most did not receive a bipolar disorder diagnosis at their initial assessment. Those with autism or intellectual disability (i.e., IQ < 70 and evidence of impairment in adaptive functioning) were exited from LAMS after their baseline assessment ($n = 10$), and therefore were not eligible for participation in the present study.

Participants were mostly male (68 %), White (70 %; African American = 14 %; multiple races = 14 %; American Indian/Alaskan Native = 2 %), and non-Hispanic (96 %), with a mean age of 9.0 years ($SD = 1.9$). Parents were mostly female (96 %), with a mean age of 38.4 years ($SD = 9.6$). Most participants (76 %; $n = 38$) received Medicaid health insurance coverage. The ChIPS and K-SADS cover 18 DSM-IV diagnoses in common, plus mania and hypomania (which can be used to rate the presence/absence of bipolar disorder) and psychosis. Therefore, 21 comparisons were involved in the main analyses. Based on the recommendations of Donner & Eliasziw (1992), a sample size of 50 individuals for these analyses provides 80 % power to detect $\kappa \geq .40$ (using $\alpha = .05$ and $\kappa = .00$ as the null hypothesis).

Measures

Kiddie-SADS-Present and Lifetime Version Plus (K-SADS-PL-W; Lingler, Bedoya, & Findling, 2007). This version of the K-SADS is updated with all mood disorder symptoms and supplemental sections covering the symptoms of pervasive developmental disorders (PDDs) added. These sections were created for LAMS and are not present in older versions of the K-SADS. Therefore, no psychometric data have been reported yet for this new PDD supplement. The K-SADS used in this study covered a total of 42 possible DSM-IV axis I diagnoses. Several items were removed from this edition of the K-SADS based on the investigator consensus at the four LAMS sites. Outdated items that referred to DSM-III diagnostic criteria were removed, as were some associated features of depression and mania (e.g., atypical and melancholic features of depression).

Children's Interview for Psychiatric Syndromes (ChIPS/P-ChIPS; Weller, et al., 1999a, 1999b). The ChIPS and P-ChIPS are described above. Both interviews were administered according to the administration manual (Rooney et al., 1999) instructions. Because duration of interviews is of interest to this study, the full introductory interview and psychosocial stressors section were administered, though data generated from these sections were not examined in the present research.

Procedures

Study personnel approached families attending their regularly-scheduled annual follow-up interviews in the LAMS study (12-, 24- or 36-months after study entry). All procedures had IRB approval. If the family agreed to participate, written informed consent and assent were obtained by the regularly-scheduled LAMS K-SADS interviewer. Seventy-one families were approached for potential participation. Of these, 50 (70.4 %) gave consent and completed data collection. Reasons for not participating included: did not wish to stay for a longer interview ($n = 15$; 75 %), child declined assent despite parent giving consent ($n = 2$; 10 %), child refused to participate in both K-SADS and ChIPS interviews ($n = 2$; 10 %), and parental informant was unable to provide consent due to not being the child's legal guardian ($n = 1$; 5 %). One family left after consenting.

ChIPS and K-SADS interviewers were graduate ($n = 7$) and postdoctoral ($n = 3$) research associates. Different interviewers conducted the K-SADS and ChIPS for each participant. All interviewers who administered the ChIPS also conducted a K-SADS interview with one or more participants in the present study. However, due to staffing changes, six participants had a K-SADS interviewer ($n = 3$) who did not administer the ChIPS to any participant. Interviewers recorded administration times and the summary diagnoses for each instrument.

All interviewers were trained to administer the ChIPS and P-ChIPS, which included a didactic component, practice interviews, and rating two previously recorded P-ChIPS interviews, as specified in the ChIPS administration manual

(Rooney et al., 1999). Inter-rater reliability was evaluated for the ChIPS by videotaping five interviews (approximately 10 % of the sample). ChIPS interviewers were required to demonstrate 100 % agreement with the original rater's diagnoses. Interviewers who failed to demonstrate inter-rater reliability were required to re-establish reliability (i.e., exact diagnostic agreement) by rating along with a live or videotaped interview before resuming independent interviews. Inter-rater reliability for ChIPS diagnoses in this study was in the "almost perfect agreement" range (mean $\kappa = .97$), using guidelines for interpreting kappa coefficients proposed by Landis and Koch (1977). K-SADS interviewers were previously trained in the K-SADS consistent with similarly rigorous training procedures as part of the LAMS study, including didactic training, observation and practice interviews, establishing reliability with live or recorded interviews prior to live administration, and regular inter-rater reliability evaluations. Therefore, all interviewers in this study had K-SADS experience (most $\geq$ 1 year) including demonstrated reliability prior to the start of this study. Conversely, only three of the interviewers (including the first author) had experience administering the ChIPS prior to the commencement of this study.

Clinician diagnoses were assigned at a case review meeting following each LAMS assessment staffed by a licensed clinician (a board-certified clinical child and adolescent psychologist) and the interviewers who administered the K-SADS and ChIPS. All information collected in the LAMS interview (e.g., K-SADS results, behavioral observations, self-report questionnaires, other interview measures of psychosocial functioning) were used by the clinician to assign LAMS study diagnoses, according to standard LAMS operating procedures. Following the completion of all LAMS study paperwork, ChIPS results were presented to the clinician. This procedure was intended to provide the clinician with as much clinical data as possible in formulating case review diagnoses. After reviewing ChIPS information and querying interviewers on any discrepancies between multiple sources of input, clinical consensus diagnoses used in this study were assigned. The order in which the ChIPS and K-SADS were administered was balanced to ensure approximately equal distribution of interview order. Twenty-five parent–child pairs received the ChIPS first, 24 parent–child pairs receive the K-SADS first, and in one parent–child pair, the parent received the K-SADS first and the child received the ChIPS first.

Results

ChIPS: K-SADS Diagnostic Agreement

Participants received an average of 3.10 ChIPS diagnoses ($SD = 1.89$) and 2.16 K-SADS diagnoses ($SD = 1.17$). The ChIPS assigned significantly more diagnoses, paired $t(49) = 4.07$, $p < .001$, similar to Swenson et al. (2007), who reported a higher mean number of diagnoses assigned by the ChIPS (4.44) than the K-SADS (3.04) in their inpatient sample.

To evaluate the first hypothesis with regard to agreement between K-SADS and ChIPS diagnoses, both overall agreement and kappa coefficients (Cohen, 1960) were calculated. Because the K-SADS integrates child and parent report, as well as other clinical observations into a summary score, K-SADS summary diagnoses were compared to diagnoses endorsed on *either* the ChIPS or P-ChIPS. As the ChIPS and K-SADS cover 18 DSM-IV diagnoses in common, plus mania, hypomania, and psychosis, 21 comparisons were made.

Six diagnoses were present in 10 % or more of the sample, allowing standard kappa coefficients to be calculated. Low-base rate (LBR) kappa coefficients (Verducci, Mack, & DeGroot, 1988) were calculated to correct for high rates of chance agreement for 11 diagnoses (see Table 1). Four diagnoses (acute stress disorder, anorexia nervosa, bulimia nervosa, and substance abuse) were not present in any participants and thus could not be analyzed. The overall mean κ was .57, with an average of 88.9 % agreement between the ChIPS and K-SADS. Using the interpretation guidelines from Landis and Koch (1977), two diagnoses are classified in the "almost perfect agreement" category (mania and enuresis). Six are classified in the "substantial agreement" category (specific phobia, social phobia, OCD, MDD, dysthymia, and encopresis). Six diagnoses (conduct disorder, separation anxiety, PTSD, hypomania, schizophrenia, and psychosis) were classified in the "moderate agreement" category, two had "fair agreement" (ADHD and GAD), and one showed "slight agreement" (ODD).

Because the ChIPS does not include instructions regarding the hierarchical rule for disruptive behavior disorders in its scoring (i.e., a diagnosis of ODD should not be assigned when CD is present), a "recoded" ODD variable was computed with this rule applied to all ChIPS interviews. This procedure, which has been used in another study (Swenson et al., 2007), provides a "common-sense" correction that clinicians would be expected to apply in routine practice. The recoded ODD variable showed *moderate* agreement with the K-SADS, compared to only *slight* agreement before applying this correction. The recoded ODD variable was included in all subsequent analyses.

Clinician: Interview Agreement

To evaluate the validity of ChIPS and K-SADS diagnoses, kappa statistics were calculated for both measures compared to clinician-assigned diagnosis (see Table 2). The

Table 1 Agreement between K-SADS and ChIPS diagnoses

Diagnosis	Positive agreements	Negative agreements	+K-SADS/−ChIPS	−K-SADS/+ChIPS	κ	Descriptive category[c]	% Agreement
ADHD	36	4	7	3	.33	Fair	80
ODD	13	12	3	22	.15	Slight	50
ODD—recoded[a]	9	16	6	19	.46	Moderate	50
Conduct disorder	2	41	0	7	.56[b]	Moderate	86
Specific phobia	8	34	2	6	.57	Moderate	84
Social phobia	2	43	1	4	.61[b]	Substantial	90
Separation anxiety	2	38	1	9	.52[b]	Moderate	80
GAD	3	40	2	5	.25	Fair	86
OCD	1	48	1	0	.74[b]	Substantial	98
PTSD	0	47	1	2	.49[b]	Moderate	94
MDD	5	39	2	4	.67	Substantial	88
Dysthymia	1	47	1	1	.65[b]	Substantial	96
Mania	1	49	0	0	1.00[b]	Almost Perfect	100
Hypomania	0	46	3	1	.48[b]	Moderate	92
Enuresis	10	40	0	0	1.00	Almost Perfect	100
Encopresis	2	46	2	0	.73[b]	Substantial	96
Schizophrenia	0	49	0	1	.50[b]	Moderate	98
Psychosis	0	47	0	3	.49[b]	Moderate	94

[a] DSM-IV hierarchical rule applied to ChIPS in cases with CD

[b] Low base rate Kappa

[c] From Landis and Koch (1977). Descriptive category: slight (κ = .00–.20); Fair (κ = .21–.40); Moderate (κ = .41–.60); Substantial (κ = .61–.80); Almost Perfect (κ = .81–1.00)

ChIPS/clinician mean kappa coefficient was .70, with a mean agreement of 93.1 %; the mean K-SADS:clinician kappa coefficient was .79, with a mean agreement of 95.4 %.

To evaluate hypothesis two, whether rates of interview-clinician agreement significantly differed between these two measures, 95 % confidence intervals (CI) were calculated for each interview's agreement with the clinician-assigned diagnoses (see Table 3). CIs overlapped for all diagnoses except ADHD (suggesting better agreement with clinician for the K-SADS than the ChIPS). However, CIs for many diagnoses were large, especially for those diagnoses where LBR kappa was computed. Even though LBR kappa tends to produce a smaller standard error estimate than the standard kappa coefficient (Verducci et al., 1988), many low-frequency diagnoses still had large standard errors because of the low number of diagnoses assigned by either instrument or the clinician. This widened the CIs and increased likelihood of the CIs overlapping.

Interview Length

To evaluate the third hypothesis, that ChIPS interviews would be shorter in time than K-SADS interviews, interview times were recorded. ChIPS interviews (child portion only) lasted an average of 35.7 min ($SD = 12.7$); P-ChIPS interviews lasted 46.2 min on average ($SD = 16.9$), with a total mean ChIPS administration time of 82.0 min ($SD = 22.9$) per participant. K-SADS interviews lasted an average of 36.0 min ($SD = 15.0$) per child and 62.4 min ($SD = 21.9$) per parent, with a mean total K-SADS time of 98.2 min ($SD = 30.9$). This difference was significant, $t(46) = 3.39$, $p = .001$, as was the parent interview difference, $t(47) = 5.30$, $p < .001$.

In addition to evaluation of the three hypotheses above, several exploratory analyses were conducted that may inform clinical utility of the ChIPS.

Impact of Interviewer Experience

To evaluate whether variation in interviewers' level of experience with clinical interviewing affected the measures' validity, each ChIPS and K-SADS was coded as to whether it was administered by an interviewer with low structured interviewing experience (≤2 years; $n = 5$) or high experience (≥3 years; $n = 5$). For both the ChIPS and K-SADS, 22 interviews were conducted by interviewers with low-experience; the remaining 28 were conducted by high-experience interviewers. Each groups' agreement with the clinician's

Table 2 Clinician Agreement with K-SADS and ChIPS Diagnoses

Diagnosis	K-SADS—clinician agreement[a]	Descriptive category[b]	% agreement	ChIPS—clinician agreement[a]	Descriptive category[b]	% Agreement
ADHD	.91	Almost Perfect	98	.37	Fair	82
ODD	.49	Moderate	76	.29	Fair	62
ODD—recoded[c]	–	–	–	.49	Moderate	74
Conduct disorder	.61[d]	Substantial	90	.87[d]	Almost Perfect	96
Specific phobia	.72	Substantial	90	.85	Almost Perfect	94
Social phobia	.78[d]	Substantial	96	.76[d]	Substantial	94
Separation anxiety	.68[d]	Substantial	92	.70[d]	Substantial	88
GAD	.44	Moderate	88	.85	Almost Perfect	96
OCD	.74[d]	Substantial	98	1.00[d]	Almost Perfect	100
PTSD	.74[d]	Substantial	98	.65[d]	Substantial	96
MDD	.90	Almost Perfect	98	.77	Substantial	94
Dysthymia	1.00[d]	Almost Perfect	100	.65[d]	Substantial	96
Mania	1.00[d]	Almost Perfect	100	1.00[d]	Almost Perfect	100
Hypomania	1.00[d]	Almost Perfect	100	.48[d]	Moderate	92
Enuresis	1.00	Almost Perfect	100	1.00	Almost Perfect	100
Encopresis	1.00[d]	Almost Perfect	100	.73[d]	Substantial	96
Schizophrenia	1.00[d]	Almost Perfect	100	.49[d]	Moderate	98
Psychosis	.49[d]	Moderate	98	.65[d]	Substantial	98

[a] Kappa coefficient

[b] From Landis and Koch (1977), Descriptive category: slight (κ = .00–.20); fair (κ = .21–.40); moderate (κ = .41–.60); substantial (κ = .61–.80); almost perfect (κ = .81–1.00)

[c] DSM-IV hierarchical rule applied to ChIPS in cases with CD

[d] LBR Kappa

diagnosis was calculated in the same manner as described above, and 95 % CIs were obtained for each kappa coefficient. Overall, CIs for low- and high-experience groups overlapped for all diagnoses for both the ChIPS and K-SADS, suggesting no significant differences between experience levels; however, given the low frequency of diagnosis for some categories, CIs for many groups were quite wide (see Table 3).

Impact of Demographic Variables

Similar exploratory analyses were conducted for child age, race/ethnicity, and sex. Participants were divided into two groups for each variable: age 7–10 ($n = 29$) vs. age 11–14 ($n = 21$), White, non-Hispanic ($n = 35$) vs. any other race and/or ethnicity ($n = 15$), and male ($n = 34$) vs. female ($n = 16$). Kappa coefficients for ChIPS-KSADS agreement were re-calculated in the same manner the primary analyses were conducted for each of these groups, and 95 % CI were obtained. Overall, results suggest that demographic variables did not affect ChIPS—K-SADS agreement. Of 54 comparisons, only two significant effects were noted (lower agreement on MDD for White, non-Hispanic participants than for all other races and ethnicities; better agreement on GAD for females than males).

Sensitivity, Specificity, Negative Predictive Value and Positive Predictive Value

To further examine the relative clinical utility of these interviews, sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) were calculated for each interview, utilizing clinician-assigned diagnoses as the "gold standard" criterion (see Table 4). These calculations must be considered exploratory because clinician diagnosis is not independent of K-SADS and ChIPS diagnoses (and is, in fact, largely dependent upon them). Nonetheless, these analyses were intended to generate hypotheses about how well both measures compare to an expert clinician who utilizes all available clinical information. The mean sensitivity estimate for the K-SADS was 72.1 % across diagnoses, and the mean specificity estimate was 99.0 %. For the ChIPS, the mean sensitivity was 79.4 %, and the mean specificity was 91.8 %.

Table 3 Confidence intervals for clinician agreement with K-SADS and ChIPS interviews

Diagnosis	K-SADS—clinician agreement (Kappa)	95 % CI	ChIPS—clinician agreement (Kappa)	95 % CI	Do CI overlap?
ADHD	.91	[0.83, 1.00]	.37	[0.21, 0.54]	No
ODD	.49	[0.36, 0.61]	.29	[0.18, 0.40]	Yes
ODD—recoded[a]	–	–	.49	[0.37, 0.61]	Yes
Conduct disorder	.61[b]	[0.22, 0.99]	.87[b]	[0.46, 1.00]	Yes
Specific phobia	.72	[0.60, 0.84]	.85	[0.77, 0.94]	Yes
Social phobia	.78[b]	[0.26, 1.00]	.76[b]	[0.33, 1.00]	Yes
Separation anxiety	.68[b]	[0.27, 1.00]	.70[b]	[0.39, 1.00]	Yes
GAD	.44	[0.26, 0.63]	.85	[0.75, 0.95]	Yes
OCD	.74[b]	[−0.09, 1.00]	1.00[b]	[−0.37, 1.00]	Yes
PTSD	.74[b]	[−0.09, 1.00]	.65[b]	[0.02, 1.00]	Yes
MDD	.90	[0.80, 1.00]	.77	[0.64, 0.89]	Yes
Dysthymia	1.00[b]	[0.04, 1.00]	.65[b]	[0.02, 1.00]	Yes
Mania	1.00[b]	[−0.37, 1.00]	1.00[b]	[−0.37, 1.00]	Yes
Hypomania	1.00[b]	[0.22, 1.00]	.48[b]	[0.02, 0.94]	Yes
Enuresis	1.00	[1.00, 1.00]	1.00	[1.00, 1.00]	Yes
Encopresis	1.00[b]	[0.33, 1.00]	.73[b]	[0.16, 1.00]	Yes
Schizophrenia	1.00[b]	N/A[c]	.49[b]	[−0.47, 1.00]	N/A[c]
Psychosis	.49[b]	[−0.47, 1.00]	.65[b]	[0.02, 1.00]	Yes
Mean	0.79		0.70		

[a] DSM-IV hierarchical rule applied to ChIPS in cases with CD

[b] LBR Kappa

[c] Not calculated due to K-SADS & clinician agreement on absence of diagnosis for all participants

Table 4 Average sensitivity, specificity, positive predictive value and negative predictive value for K-SADS and ChIPS, using clinician diagnosis as criterion

		Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
All diagnoses	K-SADS	72.1	99.0	93.4	95.0
	ChIPS	79.4	91.8	65.9	94.0
Diagnoses with ≥10 % prevalence	K-SADS	67.6	98.5	94.5	91.4
	ChIPS	91.1	85.9	76.3	90.2

Discussion

The K-SADS and ChIPS are interview measures designed to assign DSM-IV diagnoses. They differ in format (i.e., structured/semi-structured), length of interview, and amount of training and clinical judgment required by the interviewer. Both instruments have demonstrated adequate reliability and validity in a limited number of previous studies. Although the K-SADS is considered the "gold standard" for mood disorder studies and is widely used in research settings, it has several characteristics that make it almost impossible to utilize in clinical settings (Youngstrom, Findling, et al., 2005). The ChIPS is brief enough for use in clinical settings, but is used less frequently than the K-SADS in research. These measures have several common features, but the key differences between them suggested the need for a direct comparison to best determine their relative performance in assigning psychiatric diagnoses in children and adolescents.

In this study, ChIPS and K-SADS diagnoses demonstrated high rates of agreement. Seven diagnoses reached a level of "substantial" or higher agreement; an additional eight diagnoses had kappa coefficients in the range of $\kappa = .40–.59$. This suggests that the K-SADS and ChIPS performed similarly in this sample of outpatient children and adolescents with a wide range of diagnoses. Results are consistent with a prior study that compared these measures with inpatient adolescents (Swenson et al., 2007). In most cases, the ChIPS and K-SADS diagnostic interviews elicit similar data from parents and children and lead to similar clinical conclusions. In addition, ChIPS and K-SADS diagnoses do not differ significantly in rates of agreement with diagnosis assigned by a licensed clinician. The mean kappa coefficient was .79 for K-SADS/clinician and .70 for ChIPS/clinician. In other words, the expert clinician did not substantially agree with one measure more than the other, suggesting convergent validity for both measures.

There are, however, some diagnoses where ChIPS and K-SADS diverge significantly in regard to clinician agreement. For ADHD, the K-SADS agreed substantially more frequently with the clinician than the ChIPS. This appears due to both under-identification (7 cases) and over-identification (2 cases) of ADHD by the ChIPS. This unexpected result could in part reflect the fact that this is the only section on the ChIPS where the interviewer *is* prompted to exercise some clinical judgment, as the scoring instructions state “Score only if behavior occurs more often than in other children/teenagers of same age.” For conduct disorder and GAD, the ChIPS converged with clinician judgment significantly more than the K-SADS. The K-SADS under-identified both diagnoses (5 cases each) compared to the clinician and the ChIPS. This pattern of higher diagnostic identification for the ChIPS fits with the instrument’s purpose as a screening instrument. It is intended to err on the side of over- rather than under-identification of cases, to aid in quickly identifying areas in need of further investigation by the clinician, and eliminating areas not likely to be significant for a given individual.

Finally, as expected, the ChIPS interviews were shorter than K-SADS interviews, by approximately 15 min, primarily due to briefer parent interviews. There was no significant difference between K-SADS and ChIPS child interviews. Although the direction of expected difference in time was observed, the ChIPS interviews lasted about 10 min longer on average than a previously published estimate from an outpatient sample. More notably, the K-SADS interviews in this study, on average, were about 1/3 shorter than the lowest published estimate of K-SADS administration time. This difference could be due to a number of characteristics peculiar to this study. First, all participant families had been enrolled in the LAMS study for at least 12 months prior to the current data collection. This means they had previously participated in K-SADS interviews at least twice. Similarly, most interviewers had one or more years of experience administering the K-SADS as LAMS interviewers before the present study commenced. This familiarity may have significantly hastened K-SADS administration. Conversely, participants presumably had their first exposure to the ChIPS interview in this study. As mentioned above, only three of the ten interviewers had ever used this measure before this study. Although inter-rater reliability for the ChIPS was good, interviewers’ unfamiliarity with this interview may have lengthened its administration time. Generally, participants and interviewers in this study were used to the K-SADS by the time they participated in the current data collection. Thus, even though ChIPS administration was briefer than that for the K-SADS, the magnitude of this difference may have been dampened by unique features of this study.

Exploratory analyses were conducted to investigate whether characteristics of participants or interviewers affected results. Considered as a whole, it appears interviewer experience and participant age, race/ethnicity, and sex have little to no effect on the relative performance of these measures. There was an effect of race/ethnicity on agreement for MDD, and an effect of sex on agreement for GAD. While these may be real effects, results may have been found by chance, due to the sheer number of comparisons conducted in the exploratory analyses. Because of the low number of diagnoses present in many cells and wide confidence intervals present for many diagnoses, it is likely this study did not have adequate power for these additional analyses. If a larger sample or a sample with more diagnoses per participant (e.g., psychiatric inpatients) was used, these analyses may have produced different results.

Sensitivity and specificity calculations were high to very high for both measures. Sensitivity was slightly higher for the K-SADS when considering all diagnoses, but was notably higher for the ChIPS when only the “common” diagnoses (i.e., present in 10 % or more of the sample) were considered. Specificity was generally excellent for both interviews, when all diagnoses or only “common” diagnoses were considered. The negative predictive value of both interviews was always 90 % or higher in all conditions analyzed, so the absence of a diagnosis on these measures is very likely to be true. The average positive predictive value of the ChIPS for all diagnoses was 65 %, which is not much higher than chance, which suggests the ChIPS may be prone to false positives. When only the “common” diagnoses were considered, this value increased to 76 %. However, this pattern of results fits with the intended purpose of the ChIPS as a screening measure, as noted above, and should not necessarily be considered a weakness of the measure.

Similar to the one previously published comparison between these two measures, this study found significantly more diagnoses assigned by the ChIPS than the K-SADS (Swenson et al., 2007). These studies do differ in several important characteristics. Most importantly, this study used an outpatient sample but Swenson et al. (2007) studied an inpatient sample shortly after admission, who were more likely to have a greater number of diagnoses. Also, Swenson et al. (2007) studied an older sample ($M = 15.0$) with more female participants (73 %) than this study (age $M = 9.0$, 32 % females), which may also affect expected diagnostic rates.

Several limitations listed by Swenson et al. (2007) were partially addressed in the present study. The earlier study also did not include parent report, but instead compared adolescent report on the ChIPS and K-SADS. The K-SADS is not usually administered this way, as parent and child report are usually combined with interviewer judgment on this interview. Also, Swenson et al. (2007) did not compare the ChIPS and K-SADS agreement for mania, hypomania, schizophrenia, or psychosis. In addition, they excluded adolescents with developmental disabilities, psychosis, and homicidal ideation. With the exception of autism and intellectual deficiency, the remaining exclusions were not used in this study. The prior study also had two potential confounds that were not present in this study. In the earlier study, the K-SADS was always administered second (0–16 days after the ChIPS). There was also a training/experience difference between ChIPS interviewers (bachelor's-level) and K-SADS interviewers (master's- or postdoctoral-level; Swenson et al., 2007). It is possible that some of these limitations may have contributed to the lower kappa coefficients found in their study. It is important to acknowledge, however, that ChIPS interviewers were also previously trained to administer the K-SADS, which potentially influenced how interviewers administered the ChIPS.

Despite these significant methodological differences, the overall pattern of results in these two studies is similar. Table 5 summarizes the class of agreement between the two measures across the two comparison studies. The K-SADS and ChIPS showed mostly moderate or higher agreement across diagnoses, despite differences in age, sex, treatment setting, and informants between these studies. The convergence of these two studies' results is strong evidence that the ChIPS performs similarly to the K-SADS for most diagnoses.

Notably, both studies found relatively less agreement between the ChIPS and K-SADS for ADHD and GAD than most other diagnoses. It is unclear whether this pattern is the result of characteristics of either interview. As speculated above, low agreement on ADHD may be influenced by some requirement for clinical judgment in scoring this section of the ChIPS. In addition, the GAD section of the ChIPS contains two preliminary items, *both* of which need to be endorsed to move on to the remaining diagnostic criteria, whereas the GAD screen of the K-SADS contains four items, and endorsement of *any* of these will trigger completion of the GAD supplement section.

Table 5 Results of two studies comparing ChIPS and K-SADS

Class of agreement[a]	Swenson et al. (2007)	Present study
Slight	ODD[b]	None
Fair	ADHD	ADHD
	Social phobia	GAD
	GAD	
Moderate	Conduct disorder	ODD[b]
	Acute stress disorder	Conduct disorder
	PTSD	SAD
	Anorexia	PTSD
	Bulimia	Hypomania
	MDD	Schizophrenia
	Dysthymia	Psychosis
Substantial	Substance abuse	Social phobia
		Specific phobia
		OCD
		MDD
		Dysthymia
		Encopresis
Almost perfect	None	Mania
		Enuresis

[a] From Landis and Koch (1977), Descriptive category: slight (κ = .00–.20); fair (κ = .21–.40); moderate (κ = .41–.60); substantial (κ = .61–.80); almost perfect (κ = .81–1.00)

[b] DSM-IV hierarchical rule applied to ChIPS in cases with CD

A number of questions remain to be answered regarding agreement between the ChIPS and K-SADS. Most notable is the poor agreement between these interviews on diagnoses of ADHD and GAD, the only two diagnoses that demonstrated agreement below $\kappa = .40$. ADHD and GAD are common and important diagnoses in childhood mental health and low agreement between these measures on these diagnostic categories is concerning. Additional investigation into this poor performance is warranted. In the present study, the clinician diagnosis agreed more frequently with the K-SADS for ADHD diagnoses. As noted above, it appears that the ChIPS may over-diagnose ADHD in some cases and under-diagnose it in others. This may partially be due to the ChIPS intended purpose as a screening measure, but the ChIPS also under-identified ADHD in some cases. This suggests the need to adapt the ADHD section when the DSM-5 revised ChIPS is developed. For GAD, the clinician agreed more frequently with the ChIPS than the K-SADS, and it appears that the K-SADS under-diagnosed GAD.

Finally, additional questions about both measures have been identified in this study. For the K-SADS, its psychometric properties require reevaluation as existing estimates, though generally excellent, are based on small samples and versions of the K-SADS no longer in circulation. Several existing translations, modifications, or additions to this measure have not been psychometrically evaluated. This study's results appear to support the favorable evaluations of this measure, but this study was designed with the K-SADS as a benchmark, not to comprehensively evaluate this interview. A large re-validation of the most current version could help to lend evidence to the K-SADS' reputation as a valid and reliable method for assigning diagnoses. The ChIPS was independently validated in an adolescent inpatient sample (Swenson et al., 2007). That sample also included a significant number of substance abuse diagnoses, which have been absent from previous validation studies of the ChIPS. That sample, as well as the present study, also included larger proportions of racial and ethnic minorities than previous ChIPS validation samples. However, representation of Hispanics remains low in all studies. The ChIPS has not been validated by researchers not affiliated with its authors in an outpatient sample, and all previous studies of this measure have low rates of eating disorders and racial and ethnic diversity. These questions suggest the need for future validation of the ChIPS to ensure it performs consistently across diverse settings and sample populations.

Conclusion

This is the second study that has found the ChIPS to have moderate to high agreement with the K-SADS in assigning psychiatric diagnoses. Similar results were found in both investigations, despite differences in participants, interviewers, and clinical setting, suggesting these findings are reliable. This study also found the ChIPS to agree with expert clinician consensus at about the same rate as the K-SADS. As expected, the ChIPS interviews were significantly quicker to administer than the K-SADS. Researchers and clinicians must consider a multitude of factors in addition to reliability and administration time when selecting assessment tools. Other considerations include diversity of the participants/client base, supervision responsibilities for interviewers, organization of the research lab/clinic, and cost of assessment.

For researchers, the K-SADS is a frequently-used diagnostic interview for assigning psychiatric diagnoses to children and adolescents. It is comprehensive and thorough, and has excellent psychometric properties. The K-SADS is the "default" measure of choice in many studies, but its proper use requires intensive training, large amounts of clinical judgment, frequent check of inter-rater reliability, and large storage space. If, as results of this study suggest, the K-SADS and ChIPS produce similar results, researchers may choose to use the ChIPS. Interviewer training requirements would be decreased, and the time saved in assessments could decrease participant burden, or be used to administer additional measures. Interviewers may choose to supplement the ChIPS with a "drill-down" instrument to provide more fine-grained ratings of the diagnostic group being studied.

For mental health professionals in clinical settings, this suggests the ChIPS could be adopted for use in clinical assessments, especially for clinicians who wish to use EBA but have concerns about time devoted to assessment. The ChIPS is not only faster to administer, but it requires less training and judgment. In fact, it is appropriate to have a non-clinician administer the ChIPS as long as results are reviewed by a trained clinician. If this procedure were used, significant clinician time savings (and therefore financial savings) would be possible during an initial diagnostic evaluation. Somewhat higher identification of diagnoses by the ChIPS was expected, since the interview is intended to be a screening measure, and this pattern was observed. Therefore, the ChIPS may be a simple, cost-effective way to for clinicians to adopt evidence-based assessment.

Acknowledgments This study was funded by the National Institute of Mental Health (R01 MH073801) awarded to the third author and the Ohio Department of Mental Health and Addiction Services (08.1238) awarded to the first author with the third author serving as mentor.

Compliance with Ethical Standards

Conflicts of Interest Authors Matthew E. Young and Ziv E. Bell declare that they have no conflict of interest. Mary A. Fristad receives royalties from American Psychiatric Press, Inc. for the Children's Interview for Psychiatric Syndromes and royalties from Guilford Press and Child & Family Psychological Services, Inc. as well as honoraria from Physician's Post-Graduate Press and the American Occupational Therapy Association for other work.

Human and Animal Rights and Informed Consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent (assent from minors) was obtained from all individual participants included in the study.

References

Akdemir, D., & Gokler, B. (2008). Psychopathology in the children of parents with bipolar mood disorder. *Turk Psikiyatri Dergisi, 19*, 133–140.

Ambrosini, P. J. (2000). Historical development and present status of the schedule for affective disorders and schizophrenia for school-age children (K-SADS). *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 49–58.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub.

Angold, A., & Costello, J. (2000). The child and adolescent psychiatric assessment (CAPA). *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 39–48.

Angold, A., Erkanli, A., Copeland, W., Goodman, R., Fisher, P. W., & Costello, E. J. (2012). Psychiatric diagnostic interviews for children and adolescents: A comparative study. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 506–517.

Birmaher, B., Axelson, D., Strober, M., Gill, M. K., Valeri, S., Chiappetta, L., … Keller, M. (2006). Clinical course of children and adolescents with bipolar spectrum disorders. *Archives of General Psychiatry, 63,* 175–183.

Birmaher, B., Ehmann, M., Axelson, D. A., Goldstein, B. I., Monk, K., Kalas, C., … Guyer, A. (2009). Schedule for affective disorders and schizophrenia for school-age children (K-SADS-PL) for the assessment of preschool children—A preliminary psychometric study. *Journal of Psychiatric Research, 43,* 680–686.

Brasil, H. H., & Bordin, I. A. (2010). Convergent validity of K-SADS-PL by comparison with CBCL in a Portuguese speaking outpatient population. *BMC Psychiatry, 10*(83).

Chorpita, B. F., Daleiden, E. L., Ebesutani, C., Young, J., Becker, K. D., Nakamura, B. J., … Smith, R. L. (2011). Evidence-based treatments for children and adolescents: An updated review of indicators of efficacy and effectiveness. *Clinical Psychology: Science and Practice, 18*, 154–172.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37–46.

Donner, A., & Eliasziw, M. (1992). A goodness-of-fit approach to inference procedures for the kappa statistic: Confidence interval construction, significance-testing and sample size estimation. *Statistics in Medicine, 11*, 1511–1519.

Ebesutani, C., Bernstein, A., Nakamura, B. J., Chorpita, B. F., Higa-McMillan, C. K., Weisz, J. R., & Research Network on Youth Mental Health. (2010). Concurrent validity of the child behavior checklist DSM-oriented scales: Correspondence with DSM diagnoses and comparison to syndrome scales. *Journal of Psychopathology and Behavioral Assessment, 32*, 373–384.

Fristad, M. A. (2006). Psychoeducational treatment for school-aged children with bipolar disorder. *Development and Psychopathology, 18*, 1289–1306.

Fristad, M., Cummins, J., Verducci, J., Teare, M., Weller, E., & Weller, R. A. (1998a). Study IV: Concurrent validity of the DSM-IV revised children's interview for psychiatric syndromes (ChIPS). *Journal of Child and Adolescent Psychopharmacology, 8*, 227–236.

Fristad, M., Glickman, A., Verducci, J., Teare, M., Weller, E., & Weller, R. A. (1998b). Study V: Children's interview for psychiatric syndromes (ChIPS): Psychometrics in two community samples. *Journal of Child and Adolescent Psychopharmacology, 8*, 237–245.

Galanter, C. A., Hundt, S. R., Goyal, P., Le, J., & Fisher, P. W. (2012). Variability among research diagnostic interview instruments in the application of DSM-IV-TR criteria for pediatric bipolar disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 51*, 605–621.

Ghanizadeh, A., Mohammadi, M. R., & Yazdanshenas, A. (2006). Psychometric properties of the Farsi translation of the Kiddie Schedule for Affective Disorders and Schizophrenia-Present and Lifetime Version. *BMC Psychiatry*, *6*(10).

Goldstein, T. R., Birmaher, B., Axelson, D., Ryan, N. D., Strober, M. A., Gill, M. K., … Bridge, J. A. (2005). History of suicide attempts in pediatric bipolar disorder: Factors associated with increased risk. *Bipolar Disorders*, *7*, 525–535.

Hamilton, J., & Gillham, J. (1999). The K-SADS and diagnosis of major depressive disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 38*, 1065–1066.

Hunt, J. I., Dyl, J., Armstrong, L., Litvin, E., Sheeran, T., & Spirito, A. (2005). Frequency of manic symptoms and bipolar disorder in psychiatrically hospitalized adolescents using the K-SADS mania rating scale. *Journal of Child and Adolescent Psychopharmacology, 15*, 918–930.

Jensen-Doss, A. (2011). Practice involves more than treatment: How can evidence-based assessment catch up to evidence-based treatment? *Clinical Psychology: Science and Practice, 18*, 173–177.

Jensen-Doss, A., Youngstrom, E. A., Youngstrom, J. K., Feeny, N. C., & Findling, R. L. (2014). Predictors and moderators of agreement between clinical and research diagnoses for children and adolescents. *Journal of Consulting and Clinical Psychology, 82*, 1151–1162.

Kaufman, J., Birmaher, B., Brent, D., Rao, U. M. A., Flynn, C., Moreci, P., … 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 & Adolescent Psychiatry*, *36*, 980–988.

Kim, E. Y., Miklowitz, D. J., Biuckians, A., & Mullen, K. (2007). Life stress and the course of early-onset bipolar disorder. *Journal of Affective Disorders, 99*, 37–44.

Kolaitis, G., Korpa, T., Kolvin, I., & Tsiantis, J. (2003). Schedule for affective disorders and schizophrenia for school-age children-present episode (K-SADS-P): A pilot inter-rater reliability study for Greek children and adolescents. *European Psychiatry: The Journal of the Association of European Psychiatrists, 18*, 374–375.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.

Lauth, B., Arnkelsson, G. B., Magnússon, P., Skarphéðinsson, G. Á., Ferrari, P., & Pétursson, H. (2010). Validity of K-SADS-PL (Schedule for Affective Disorders and Schizophrenia for School-Age Children—Present and Lifetime Version) depression diagnoses in an adolescent clinical population. *Nordic Journal of Psychiatry, 64*, 409–420.

Leffler, J. M., Riebel, J., & Hughes, H. M. (2015). A review of child and adolescent diagnostic interviews for clinical practitioners. *Assessment, 22*, 690–703.

Lewinsohn, P. M., Shankman, S. A., Gau, J. M., & Klein, D. N. (2004). The prevalence and comorbidity of subthreshold psychiatric conditions. *Psychological Medicine, 34*, 613–622.

Lingler, J., Bedoya, D., & Findling, R. L. (2007, January). *The Kiddie-SADS Present and Lifetime Version Plus (K-SADS-PL-W).* [Unpublished measure] Department of Psychiatry, University Hospitals of Cleveland, Case Western Reserve University, Cleveland, OH.

Luca, P., Laurin, N., Misener, V. L., Wigg, K. G., Anderson, B., Cate-Carter, T., … Barr, C. L. (2007). Association of the dopamine receptor D1 gene, DRD1, with inattention symptoms in families selected for reading problems. *Molecular Psychiatry, 12*, 776–785.

Mattai, A. A., Tossell, J., Greenstein, D. K., Addington, A., Clasen, L. S., Gornick, M. C., et al. (2006). Sleep disturbances in childhood-onset schizophrenia. *Schizophrenia Research, 86*, 123–129.

Nottelmann, E. (2001). National Institute of Mental Health research roundtable on prepubertal bipolar disorder. *Journal of the American Academy of Child and Adolescent Psychiatry, 40*, 871–878.

Paing, W. W., Weller, R. A., Dixon, T. A., & Weller, E. B. (2010). Face-to-face versus telephone administration of the parent's version of the children's interview for psychiatric syndromes (P-ChIPS). *Current Psychiatry Reports, 12*, 122–126.

Puig-Antich, J., & Chambers, W. (1978). *The schedule for affective disorders and schizophrenia for school-age children (Kiddie-SADS)*. New York: New York State Psychiatric Institute.

Puig-Antich, J., & Ryan, N. (1986). *The schedule for affective disorders and schizophrenia for school-age children (Kiddie-SADS)-1986*. Pittsburgh: Western Psychiatric Institute and Clinic.

Quintana, H., Butterbaugh, G., Purnell, W., & Layman, A. K. (2007). Fluoxetine monotherapy in attention-deficit/hyperactivity disorder and comorbid non-bipolar mood disorders in children and adolescents. *Child Psychiatry and Human Development, 37*, 241–253.

Reich, W. (2000). Diagnostic interview for children and adolescents (DICA). *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 59–66.

Renou, S., Hergueta, T., Flament, M., Mouren-Simeoni, M., & Lecrubier, Y. (2004). Entretiens diagnostiques structures en psychiatrie de l'enfant et de l'adolescent [Diagnostic structured interviews in child and adolescent's psychiatry]. *L'Encephale, 30*, 122–134.

Rolon-Arroyo, B., Arnold, D. H., Harvey, E. A., & Marshall, N. (2016). Assessing attention and disruptive behavior symptoms in preschool-age children: The utility of the Diagnostic Interview Schedule for Children. *Journal of Child and Family Studies, 25*, 65–76.

Rooney, M. T., Fristad, M. A., Weller, E. B., & Weller, R. A. (1999). *Administration manual for the Children's Interview for Psychiatric Syndromes (ChIPS)*. Washington, D.C.: American Psychiatric Press, Inc.

Schoeman, R., Carey, P., & Seedat, S. (2009). Trauma and posttraumatic stress disorder in South African adolescents: a case-control study of cognitive deficits. *The Journal of Nervous and Mental Disease, 197*, 244–250.

Shaffer, D., Fisher, P., Lucas, C., Dulcan, M., & Schwab-Stone, M. E. (2000). NIMH diagnostic interview schedule for children version IV (NIMH DISC-IV): Description, differences from previous versions, and reliability of some common diagnoses. *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 28–38.

Shanee, N., Apter, A., & Weizman, A. (1997). Psychometric properties of the K-SADS-PL in an Israeli adolescent clinical population. *Israel Journal of Psychiatry and Related Sciences, 34*, 179–186.

Sheehan, D. V., Sheehan, K. H., Shytle, R. D., Janavs, J., Bannon, Y., Rogers, J. E., … Wilkinson, B. (2010). Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID). *Journal of Clinical Psychiatry, 71*, 313–326.

Swenson, L. P., Esposito-Smythers, C., Hunt, J. I., Hollander, B. L. G., Dyl, J., Rizzo, C. J., et al. (2007). Validation of the children's interview for psychiatric syndromes (ChIPS) with psychiatrically hospitalized adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry, 46*, 1482–1490.

Teare, M., Fristad, M., Weller, E., Weller, R., & Salmon, P. (1998a). Study I: Development and criterion validity of the children's interview for psychiatric syndromes (ChIPS). *Journal of Child and Adolescent Psychopharmacology, 8*, 205–211.

Teare, M., Fristad, M., Weller, E., Weller, R., & Salmon, P. (1998b). Study II: Concurrent validity of the DSM-III-R children's interview for psychiatric syndromes (ChIPS). *Journal of Child and Adolescent Psychopharmacology, 8*, 213–219.

Van Meter, A. R., Moreira, A. L. R., & Youngstrom, E. A. (2011). Meta-analysis of epidemiologic studies of pediatric bipolar disorder. *The Journal of Clinical Psychiatry, 72*, 1–478.

Verducci, J. S., Mack, M. E., & DeGroot, M. H. (1988). Estimating multiple rater agreement for a rare diagnosis. *Journal of Multivariate Analysis, 27*, 512–535.

Weller, E., Weller, R., Fristad, M., Rooney, M., & Schecter, J. (2000). Children's interview for psychiatric syndromes (ChIPS). *Journal of the American Academy of Child and Adolescent Psychiatry, 39*, 76–84.

Weller, E. B., Weller, R. A., Rooney, M. T., & Fristad, M. A. (1999a). *Children's Interview for Psychiatric Syndromes (ChIPS)*. Washington, D.C.: American Psychiatric Press, Inc.

Weller, E. B., Weller, R. A., Rooney, M. T., & Fristad, M. A. (1999b). *Children's Interview for psychiatric syndromes—Parent version (P-ChIPS)*. Washington, D.C.: American Psychiatric Press, Inc.

Youngstrom, E. A. (2013). Future directions in psychological assessment: Combining evidence-based medicine innovations with psychology's historical strengths to enhance utility. *Journal of Clinical Child & Adolescent Psychology, 42*, 139–159.

Youngstrom, E. A., Findling, R. L., Youngstrom, J. K., & Calabrese, J. R. (2005). Toward an evidence-based assessment of pediatric bipolar disorder. *Journal of Clinical Child and Adolescent Psychology, 34*, 433–448.