

Diagnosing Oppositional Defiant Disorder (ODD) Using the Anxiety Disorders
Interview Schedule for *DSM-IV*: Parent Version (ADIS-P)

Scott Robert Anderson

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of
Master of Science
In
Psychology

Thomas H. Ollendick
Russell T. Jones
Jungmeen Kim

12/01/2009
Blacksburg, VA

Keywords: Oppositional Defiant Disorder, Interview Schedule, Disruptive Behaviors,
Assessment

Copyright 2009, Scott R. Anderson

Diagnosing Oppositional Defiant Disorder (ODD) Using the Anxiety Disorders
Interview Schedule for *DSM-IV*: Parent Version (ADIS-P)

Scott Robert Anderson

ABSTRACT

The purpose of this study was to determine whether the Anxiety Disorders Interview Schedule for *DSM-IV*: Parent Version (ADIS-P) is a valid diagnostic tool in assessing Oppositional Defiant Disorder (ODD) in youth. Although there is considerable evidence that the ADIS-P is effective when diagnosing anxiety disorders in youth, no studies have yet examined its utility in assessing ODD, even though the ADIS-P contains an ODD module. In contrast, a number of studies support the Diagnostic Interview Schedule for Children-Version IV (DISC-IV) as a reliable and valid tool for assessing ODD. The two diagnostic interviews have not been compared to determine whether the ADIS-P might be equally valid to the DISC-IV in diagnosing ODD. In this study, the ADIS-P and DISC-IV ODD modules were administered in a counterbalanced order to the parents of a clinical sample of 53 children between 8 and 13 years of age referred for the treatment of ODD. It was hypothesized that the ODD module of the ADIS-P would be reliable, as evidenced by inter-rater correspondence, and valid as determined by its concurrent validity with the DISC-IV and its relations with the Behavior Assessment System for Children (BASC) Aggression and Conduct Problems scales as well as the Disruptive Behavior Disorders rating scale (DBD). Both of these latter instruments were completed by parents and teachers of the referred youth. Results suggest that the ADIS-P provides a valid assessment of ODD, giving clinicians and researchers another empirically-supported interview to use when assessing children's disruptive behaviors.

Acknowledgments

I would like to express my gratitude to my advisor, Dr. Tom H. Ollendick, and to my committee members, Dr. Russell T. Jones and Dr. Jungmeen Kim, for their support and guidance in this process. I would also like to thank Dr. Matthew S. Fritz for providing an exceptional education in advanced statistical analyses.

I am also thankful for the love and support that I receive from my wife, Hannah, and my two daughters, Emmaleigh and Isabelle. Their smiling faces when I (finally) come home each day lift my spirits. Their sacrifices and encouragement motivate me to succeed. I love you.

Finally, I must acknowledge my brother, Craig, and my sister, Bethany, for wonderful friendship, great advice, and so many good times over the years. I especially want to thank my father, Dr. Robert W. Anderson, and my mother, Ruth V. Anderson, for the personal and professional values and direction that they have instilled in me.

Table of Contents

Acknowledgments.....	iii
List of Tables.....	v
Introduction.....	1
Diagnostic Interview Schedule for Children–Version IV (DISC–IV).....	1
Anxiety Disorders Interview Schedule for <i>DSM–IV</i> : Parent Version (ADIS–P).....	3
Hypotheses.....	6
Method.....	6
Participants.....	6
Materials.....	7
Procedure.....	9
Results.....	9
Testing for Order Effects.....	9
Agreement between the ADIS–P ODD Module and the DISC–IV ODD Module.....	10
Comparing Diagnostic Interviews with the DBD and the BASC.....	11
Discussion.....	12
Limitations.....	15
Clinical Implications.....	16
References.....	17
Appendix—Revised DBD.....	30

List of Tables

Table 1. Descriptive Statistics and t-Tests.....	22
Table 2. Rates of Agreement and Kappa Coefficients for ADIS–P and DISC–IV Symptoms (N = 53).....	23
Table 3. Rates of Agreement and Kappa Coefficients for ADIS–P and DISC–IV Diagnostic Decisions (N = 53).....	24
Table 4. Correlations among Maternal-Report Questionnaire Scales, Interference Ratings, Interview Symptom Counts, and Clinician Severity Ratings (n = 52).....	25
Table 5. Correlations among Paternal-Report Questionnaire Scales, Interference Ratings, Interview Symptom Counts, and Clinician Severity Ratings (n = 46).....	26
Table 6. Correlations among Teacher-Report Questionnaire Scales, Interference Ratings, Interview Symptom Counts, and Clinician Severity Ratings (n = 31).....	27
Table 7. t-Tests for ADIS–P and DISC–IV Correlations.....	28
Table 8. Correlations among Adult Reporters on Questionnaire Scales.....	29

Introduction

The study of child psychopathology and the clinical treatment of children who experience various forms of psychopathology necessitate the use of psychometrically sound assessments to establish reliable and valid diagnoses. In recent years, third-party payers have emphasized the importance of formal diagnoses in clinical settings as part of an effort to keep health care costs manageable (Achenbach, 1998). At the same time, research interest in childhood disorders has grown, leading to an increase in the use of diagnostic tools designed to assess these psychological disorders in children (see Schniering, Hudson, & Rapee, 2000). Furthermore, research seeking to provide empirical support to various psychosocial treatments for childhood psychopathologies requires the use of evidence-based diagnostic measures. The dramatic growth of research in this field is encouraging (Silverman & Hinshaw, 2008) and contributes to the increasing popularity of structured and semi-structured diagnostic interviews.

Inasmuch as children are often referred to psychological treatment because of behavioral and conduct problems, this study will focus on the assessment of these problems in referred youth. Previous reviews (McMahon & Frick, 2005) have recommended that assessments of children with conduct problems include information gathered from several sources using different methods, including diagnostic interviews whenever possible. This is particularly important because interviews allow the clinician to obtain clinically significant qualitative information that might not be captured in traditional clinical interviews, self-report questionnaires, and parent and teacher rating measures. Thus, there is a need for diagnostic interviews that can be used in both research and clinical settings.

Diagnostic Interview Schedule for Children–Version IV (DISC–IV)

Recently, much research with conduct problem children and adolescents has focused on the Diagnostic Interview Schedule for Children–Version IV (DISC–IV; Shaffer, Fisher, Lucas, Dulcan, & Schwab-Stone, 2000), a comprehensive, structured diagnostic instrument based on the *Diagnostic and Statistical Manual–IV (DSM–IV; American Psychiatric Association, 2000)*. The interviewee provides yes/no answers to questions regarding symptoms not only of conduct problems but also of most psychological disorders observed in children and adolescents. Questions also explore how much the endorsed symptoms interfere in the child's life. Diagnoses are then derived from a structured algorithm that includes the symptom counts, interference levels, and other relevant *DSM–IV* criteria. As such, the DISC–IV is now one of the most

extensively studied diagnostic interviews in child psychopathology research and is frequently used as a gold standard for providing concurrent validity to other diagnostic measures (e.g., Dewey, Kaplan, Crawford, & Fisher, 2001; Hayes, McReynolds, & Wasserman, 2005; Lemery-Chalfant et al., 2007). Moreover, the DISC-IV is often used to estimate the prevalence of childhood psychopathology in community settings (e.g., Roberts, Roberts, & Xing, 2007) and to identify child participants in both prevention and treatment studies (e.g., Dewey et al., 2001; Nemeroff et al., 2008). The DISC-IV has been shown to possess acceptable test-retest reliability (Shaffer et al., 2000), and the Oppositional Defiant Disorder (ODD) module has external evidence of its validity (Friman et al., 2000). Indeed, the DISC-IV is the most commonly used diagnostic tool in studies of treatments for conduct disorders among children and adolescents (Frick & McMahon, 2008).

Although the DISC-IV is an effective diagnostic tool for assessing children and adolescents, it does have drawbacks compared to other available diagnostic interviews. The DISC-IV is highly *structured* and was designed to be administered by interviewers with minimal training. Although the structured format of the DISC-IV is an asset for encouraging high levels of reliability in lay persons, it can also limit the amount of clinical judgment that is rendered in the diagnostic process and potentially affect the validity of the diagnostic decisions that are subsequently made. As noted by Grills and Ollendick (2002), the high level of structure could also hinder the clinician from developing rapport with the interviewee. Without strong rapport with the clinician, the parent or child might not feel comfortable endorsing relevant symptoms. Similarly, inasmuch as the DISC-IV is designed to be administered by individuals who do not always have the requisite clinical training to gather important qualitative information, important details may be lacking. For example, although the DISC-IV only requires the client to endorse the presence and frequency of a behavior (e.g., arguing with adults at least once a week), it often is helpful for the clinician to obtain more details (e.g., with which adult the child typically argues, what the arguments are about, what has led up to the arguments, in what situations do they occur, etc.).

It is also noteworthy that the DISC-IV alters the phrasing and presentation of some *DSM-IV* symptoms. In the ODD module of the DISC-IV, for example, three of the eight *DSM-IV* symptoms of ODD are subdivided; in other words, one symptom is split in half, and each half is presented to the interviewee separately. An example of this is the *DSM-IV* symptom of defiance,

which the DISC-IV splits into two questions: doing what one is asked not to do, and not doing what one is asked to do. Although this might provide important specific knowledge about a child's problematic behaviors and symptoms, it might also influence the number of symptoms that parents endorse, because they are given two opportunities to endorse the same symptom. Furthermore, the DISC-IV counts a symptom of ODD as present if the parent indicates the behavior has been occurring at a frequency of at least once a week for a period of at least six months. However, it is unclear how the particular criterion for frequency of "at least once a week" was determined. The *DSM-IV* states that behaviors symptomatic of ODD meet criteria if they occur "often"; thus, there is room for debate about whether "once a week" is "often" for these behaviors, or even whether it is "too often" and pathognomic of disorder.

Furthermore, although the ODD module of the DISC-IV includes six items that assess for impairment related to the symptoms of ODD, the diagnostic algorithm does not specify how these items should be used when making diagnostic decisions (i.e., there is no consensus regarding what would indicate a clinical level of impairment). These items ask the parent whether the child's behaviors cause his or her caretakers (i.e., parents and/or step-parents) to be upset, prevent the child from doing things with the family, keep the child from participating in activities with peers, cause the child to feel upset, create difficulties with his or her schoolwork, and cause the child's teacher to be upset. The standard algorithm determines diagnoses based solely on the number of symptoms that the parent endorsed as occurring once a week for 6 months or longer, without integrating information from the impairment items; however, the scoring instructions for the DISC-IV provide four suggestions for how to supplement the standard scoring algorithm by including impairment data. To meet the impairment criterion, Impairment Algorithm A requires at least one of the impairment items to be endorsed by the parent at a moderate or severe level (2 or 3 on 0-3 scale). Impairment Algorithm B requires at least two of the impairment items to be endorsed at a moderate or severe level. Impairment Algorithm C requires at least one impairment item to be endorsed at a severe level. The impairment criterion is met in Impairment Algorithm D if criteria for either Impairment Algorithms B or C are met.

Anxiety Disorders Interview Schedule for DSM-IV: Parent Version (ADIS-P)

An alternative diagnostic interview that has also received wide use in the diagnosis of diverse forms of child psychopathology is the Anxiety Disorders Interview Schedule for *DSM-*

IV: Child and Parent Versions (ADIS–C/P; Silverman & Albano, 1996). It was developed to follow *DSM–IV* guidelines and to provide reliable and valid diagnoses while allowing for clinician input into the process. As such, the ADIS–C/P is a *semi-structured* interview that allows for and requires considerable clinical judgment to be administered. Rather than asking to quantify specifically the frequency of behaviors, parents and children are typically asked to endorse the presence of symptoms and to give subjective ratings regarding severity (e.g., how much anxiety the child feels in specific situations or how problematic certain behaviors are in certain settings). Interviewees are also asked to provide estimates for how distressing or interfering each endorsed symptom is for the child. The ADIS–C/P is the most common diagnostic interview in studies that assess and treat childhood anxiety disorders (Silverman & Ollendick, 2005, 2008; Silverman, Pina, & Viswesvaran, 2008). However, these studies have been limited mostly to the anxiety disorders section of the ADIS–C/P; little research has been done to examine the validity of the ADIS–C/P modules that assess for other diagnoses, with one notable exception (Jarrett, Wolff, & Ollendick, 2007). Jarrett et al. (2007) found that the ADIS–C/P module that assesses symptoms of Attention-Deficit/Hyperactivity Disorder had acceptable concurrent and predictive validity. However, no studies have been published regarding the validity and reliability of the ODD module of the ADIS–P (the child version does not include an ODD module). As a result, this diagnostic instrument is not often used in clinical or research settings when other diagnoses (such as ODD) are of primary interest.

There is reason to believe that the ADIS–P will be effective at assessing ODD symptoms. Although it does not instruct the assessor to ask for a frequency estimate for each behavior endorsed by the parent, it does ask parents to endorse those symptoms the child displays “a lot more than would be appropriate for his or her age.” Thus, parents are encouraged to report only those behaviors that they perceive to occur “often” but also that are developmentally inappropriate, which can help prevent over-diagnosing the disorder when parents think that a child’s behavior is within normative limits or expectations.

Because this study focuses on the validity of the ODD module of the ADIS–P, it is also important to address the issue of reliability. If reliability is unacceptably low, then validity cannot be meaningfully examined. Previous research has indicated that the ADIS–C/P possesses acceptable test-retest reliability, including the ODD module at the diagnostic level ($\kappa = .62$; Silverman, Saavedra, & Pina, 2001, p. 941). Diagnoses derived from the ADIS–C/P have good to

excellent interrater agreement regardless of whether the diagnoses are based on the parent's or child's report, except for ADHD diagnoses based solely on the child's interview (Lyneham, Abbott, & Rapee, 2007; also see Jarrett et al., 2007, regarding ADHD). The portions of the ADIS-C/P that assess anxiety disorders have received both concurrent and discriminant evidence of validity (Wood, Piacentini, Bergman, McCracken, & Barrios, 2002; Grills-Taquechel & Ollendick, 2008). Furthermore, Grills and Ollendick (2003) examined agreement among informants, as well as between informants and diagnoses developed in a clinical consensus meeting. Although parent-child agreement and child-consensus agreement were both reported to be rather poor for all diagnostic categories, parent-consensus agreement ranged from poor to good, with only one diagnosis (i.e., Specific Phobia) in the poor range. Parent-consensus agreement for ODD diagnoses was good ($\kappa = .64$; Grills & Ollendick, 2003, p. 34); however, little research has compared the ADIS-P ODD module with other measures of ODD, and few treatment studies have used the ADIS-P to assess ODD (for an exception see Costin & Chambers, 2007). Thus, although research indicates that the ODD module of the ADIS-P has acceptable test-retest reliability similar to the DISC-IV ($\kappa = .62$ and $\kappa = .59$, respectively; Silverman et al., 2001, p. 941; Shaffer et al., 2000, p. 34), research has yet to test the validity of this module. If the ODD module of the ADIS-P yields similar evidence of validity as the DISC-IV ODD module, then the ADIS-P would be an acceptable alternative to the DISC-IV for assessing ODD. Furthermore, the test-retest estimates and evidence of validity for the anxiety modules of the ADIS-P suggest that the ADIS-P is an excellent choice when assessing comorbid anxiety disorders in these populations (Silverman et al., 2001; Wood et al., 2002).

Although both the ADIS-C/P and the DISC-IV have similarly low parent-child agreement rates (Grills & Ollendick, 2002; Jensen et al., 1999), the ADIS-C/P has higher test-retest reliability coefficients for Separation Anxiety Disorder, Social Phobia, and Generalized Anxiety Disorder than the DISC-IV (Silverman et al., 2001), suggesting that it might be a better choice when these disorders are of interest. Perhaps most importantly, the anxiety modules of the ADIS-C/P have strong evidence for validity and reliability, whereas some research indicates that anxiety as reported on the DISC-IV might be different from the anxiety identified by other diagnostic instruments (MTA Cooperative Group, 2000). From a clinician's perspective, if both ODD and an Anxiety Disorder are suspect and if the clinician is to use a diagnostic interview, it

would make sense for the clinician to use one interview that reliably and validly assesses both disorders rather than two separate ones.

Hypotheses

The present study addressed some of these shortcomings. In order to provide further support for the validity of the ODD module of the ADIS–P, this module and the ODD portion of the DISC–IV were administered to the parents of youth ages 8–13 years old. The diagnoses derived from these interviews were compared. It was hypothesized that DISC–IV diagnoses determined using Impairment Algorithm D (i.e., the impairment criterion is met if the parent endorses at least one impairment item at a severe level or endorses at least two items at a moderate level) would yield agreement rates with ADIS–P diagnoses. Furthermore, results from both interview modules were compared with the Aggression and Conduct Problems scales of the Behavior Assessment System for Children (BASC; Reynolds & Kamphaus, 1992) and the ODD section of the Disruptive Behavior Disorders rating scale (DBD; Barkley, 1997; Pelham, Gnagy, Greenslade, & Milich, 1992). Inasmuch as questionnaires are often used when time and financial resources are limited, they are important tools in the diagnostic process. Furthermore, the BASC and the DBD have satisfactory psychometric properties and are frequently used in clinical and research settings. It has even been suggested that behavior rating scales (e.g., the DBD) might be at least as valid as parent interviews when assessing ADHD in children (Wright, Waschbusch, & Frankland, 2007), although this question has not been directly examined in relation to ODD.

Thus, this study attempted to provide crucial information in several domains. It was hypothesized that the results would provide support for the convergent validity of the ODD module of the ADIS–P, giving researchers and clinicians another empirically supported diagnostic tool with which to evaluate children. If the hypotheses were supported, the ADIS–P could be particularly helpful when assessing oppositional youth when comorbid anxiety disorders are suspected, because research has yet to determine the full importance of comorbid anxiety disorders in this population (Chase & Eyberg, 2008).

Method

Participants

Participants were drawn from a larger randomized clinical control trial of youth with ODD (“Mediators, Moderators, and Treatment Efficacy of Two Forms of Psychosocial Treatment for Oppositional Children,” Ollendick, 2007) and came from the New River Valley

and Roanoke Valley metropolitan areas of southwestern Virginia. Children and families were excluded from participation if the child met criteria for a pervasive developmental disorder, had an estimated IQ below 80, or failed to meet criteria for a diagnosis of ODD (primary or secondary). Children ($N = 53$) were 8–13 years of age ($Mean = 9.74$, $SD = 1.91$) and were referred for treatment. Nineteen of the children (36%) were female. Forty-six (87%) children were identified as Caucasian, 4 (8%) as African American, and 3 (6%) as Asian-American. Participants in this study were the parents or guardians of these children. Thirty-five (66%) of the interviews were conducted with the child's mother or female guardian; the other interviews were conducted with both parents or guardians. If one parent was absent from the assessment session, then he or she was asked to complete the forms via mail. In addition, 31 teachers (58%) participated by completing questionnaires via mail.

Materials

Diagnostic Interview Schedule for Children–Version IV (DISC–IV). The DISC–IV was used as the criterion measure or “gold standard” for ODD diagnoses. The hardcopy, parent version of the ODD section of the DISC–IV was administered to the parents or guardians of the children. The parent version of the DISC–IV can be administered to parents of children between the ages of 6 and 17 years old. The DISC–IV provides a tally of endorsed symptoms, as well as frequency estimates as endorsed by the parents. In addition, the DISC–IV also includes items that ask parents to determine how much the child's symptoms interfere with family activities, friendships, and the child's academic performance on a 4-point, 0 (no impairment) to 3 (severe impairment) rating scale. Scoring the DISC–IV and assigning diagnoses are accomplished by following guidelines provided by the diagnostic algorithm, which incorporates the number of symptoms endorsed with the duration and frequency of the symptoms. For purposes of this study, parents' responses on the impairment items were summed (possible range of scores = 0–18) in order to calculate correlations with other measures. It is important to note that the standard algorithm does not utilize the impairment items; however, the instructions for scoring the DISC–IV include suggestions for different ways to include this additional impairment information. In order to examine how impairment endorsed on the DISC–IV might affect agreement between the ADIS–P and the DISC–IV, DISC–IV diagnoses were derived according to each of the four impairment algorithms described above, in addition to the standard algorithm, which does not incorporate information regarding impairment.

Anxiety Disorders Interview Schedule for DSM-IV: Child and Parent Versions (ADIS-C/P). Parallel versions of the ADIS-C/P are typically administered to parents and their children between the ages of 7 and 16 years old. However, the child version of the ADIS-C/P does not include an ODD module, so only the parent version was used in the present study. The ODD module of the ADIS-P includes an initial screening question, followed by items related directly to the eight *DSM-IV* symptoms of ODD. The ADIS-P requests the parents to indicate the severity of these symptoms on a 9-point scale (0–8, with a rating ≥ 4 suggesting a clinical level of interference). Scoring the ADIS-P and assigning diagnoses involves following *DSM-IV* criteria as outlined in the ADIS-P while exercising clinical judgment. Clinician severity ratings, based on information provided by the parents, are also determined on a 9-point scale (0–8).

Behavior Assessment System for Children-2nd Edition (BASC). The BASC (Reynolds & Kamphaus, 1992) is a coordinated system of instruments that evaluates the behaviors, thoughts, and emotions of children and adolescents. This system allows clinicians to gather data from three types of informants: the child, classroom teachers, and parents. For purposes of the present study, the Aggression and Conduct Problems scales of the Parent Rating Scales and Teacher Ratings Scales were enlisted. Average internal consistency for the Teacher Ratings Scales is .80, and median test-retest reliability is .91 (Kamphaus & Frick, 2005). The Parent Rating Scales also have good to excellent internal consistency (.70s–.90s; Kamphaus & Frick, 2005) and test-retest reliability over a 2- to 8-week period (.74–.94; Reynolds & Kamphaus, 1992). Research also provides evidence that the BASC demonstrates good convergent and DISC-IV discriminant validity (Merrell, Blade, Lund, & Kempf, 2003; Merydith, 2001), as well as acceptable criterion validity (Reynolds & Kamphaus, 1992). In this sample, the internal consistency (measured using Cronbach's α) of the Aggression scale was .83 for mothers, .86 for fathers, and .88 for teachers. The Conduct Problems scale also demonstrated good internal consistency for mothers (Cronbach's $\alpha = .78$), fathers (Cronbach's $\alpha = .87$), and teachers (Cronbach's $\alpha = .84$).

Disruptive Behavior Disorders Rating Scales (DBD). The DBD (Barkley, 1997; Pelham et al., 1992) is derived from the *DSM-IV* symptoms for ADHD, ODD, and Conduct Disorder (CD). For the present study, the Barkley (1997) revision of the scale was used; ratings of ODD symptoms from teachers and parents were obtained (see Appendix for a copy of these items). Respondents rate each symptom on a 4-point scale ranging from 0 (not at all) to 3 (very much); for purposes of this study, items were recoded to indicate presence or absence of the symptom

following Barkley's guidelines (i.e., responses of 0 or 1 indicated the absence of a symptom, whereas responses of 2 or 3 indicated the presence of a symptom). The DBD has been shown to have good psychometric properties (Pelham et al., 1992) and to have supportive evidence of validity (Wright et al., 2007). In this sample, internal consistency was calculated for the 8 items that assess symptoms of ODD. The eight-item subscale of the DBD had acceptable internal consistency for mothers (Cronbach's $\alpha = .80$), fathers (Cronbach's $\alpha = .88$), and teachers (Cronbach's $\alpha = .92$).

Procedure

Assessors administered the complete ADIS-P and the ODD section of the DISC-IV to the parents during a 2-hour assessment session; for purposes of this study, only the ODD modules of each interview schedule were studied further. In order to minimize possible ordering effects, the diagnostic measures were administered in a counterbalanced order. After administering the interview schedules, assessors assigned diagnoses obtained from each measure, based on the guidelines set forth in the respective manuals and algorithms. The BASC and the DBD were completed by the parents during this assessment session. If one parent was unable to attend the assessment session, then the questionnaires were sent home with a pre-addressed envelope, and the absent parent (usually the father or male guardian) was asked to mail the completed forms to the Child Study Center. The teacher forms were mailed to teachers identified by the parents as having the most up-to-date information about their child. Typically, these teachers were homeroom teachers.

To ensure that diagnostic protocols were followed and administered in a competent manner, an expert rater (Ph.D.-level clinician) checked at least 20% of the diagnoses by reviewing videotapes of the assessment sessions. ADIS-P diagnoses of ODD had perfect agreement ($\kappa = 1.00$); reliability statistics for DISC-IV diagnoses of ODD were not available for the present study but are presently being determined.

Results

Testing for Order Effects

Means and standard deviations for variables are presented in Table 1. A series of *t*-tests were conducted in order to determine whether the order in which the interviews were administered influenced the results. First, a paired samples *t*-test compared the number of ODD symptoms endorsed on the ADIS-P with the number of symptoms endorsed on the DISC-IV.

The results indicated that the difference in the means was not significant, $t(52) = 1.49, p = .14$. Additional independent samples t -tests indicated that means for participants who were administered the ADIS-P first did not differ significantly from those of participants who were administered the DISC-IV first (see Table 1).

Agreement between the ADIS-P ODD Module and the DISC-IV ODD Module

The total number of symptoms endorsed on the ADIS was correlated with the total number of symptoms endorsed on the DISC, $r = .64, p < .001$. Next, rates of agreement and kappa coefficients were determined for each of the eight ODD symptoms to provide more information about the agreement between the ADIS-P and the DISC-IV (see Table 2). Overall agreement rates ranged from 57% to 96%. Based on guidelines proposed by Landis and Koch (1977), most of the kappa coefficients were in the moderate range (.41–.60). However, two of the kappa coefficients were lower: the kappa coefficient for agreement for the symptom assessing how often children are angry was in the slight range (.00–.20) and the kappa for the temper symptom was in the fair range (.21–.40). A kappa coefficient could not be calculated for the symptom that assesses the symptom of arguing, even though overall agreement was 96%. Cicchetti and Feinstein (1990) have pointed out that the kappa coefficient can be low when the distribution of agreement is skewed, even when overall agreement is actually very high. In cases where more in-depth information about agreement is needed, Cicchetti and Feinstein suggest that positive agreement (in this case, agreeing that a child does exhibit the symptom) and negative agreement (i.e., agreeing that a child does not display the symptom) rates can be more helpful than simply looking at total agreement or the kappa statistic (see Table 2). Because the sample used in this study was a clinical sample of youth referred for treatment of ODD, there were very few cases that were not diagnosed with ODD: Forty-nine children (92%) were diagnosed based on the ADIS-P, and 49 children (92%) were diagnosed based on the DISC-IV using Impairment Algorithm D. (50 children were assigned DISC-IV diagnoses when the impairment criterion was not included in the algorithm.)

The rates of positive and negative agreement for the endorsement of each symptom on the ADIS-P and the DISC-IV suggest that the majority of children in the sample were positive for the first three symptoms of ODD (i.e., losing temper frequently, arguing often, and frequently being defiant) on both diagnostic interviews. For these symptoms, there was less opportunity to examine the negative agreement of the DISC-IV and the ADIS-P, because so few parents did

not endorse these items. For the other five symptoms, rates of negative agreement ranged from 47% to 86%. Positive agreement was very high ($\geq 93\%$) for the first five symptoms; however, positive agreement was lower (55%–78%) for the last three symptoms (see Table 2).

Rates of agreement and kappa coefficients were also calculated to examine overall diagnostic agreement between the ADIS–P and the DISC–IV (see Table 3). DISC–IV diagnostic decisions were determined using four different algorithms for meeting the impairment criterion and the one algorithm that did not include the impairment criterion. The highest rates of negative and overall agreement and the highest kappa coefficient were obtained when the ADIS–P diagnostic decisions were compared with DISC–IV diagnoses derived using Impairment Algorithm D. The highest positive agreement was obtained when Impairment Algorithm C was used to determine DISC–IV diagnoses; however, this algorithm led to the lowest rate of negative agreement.

Comparing Diagnostic Interviews with the DBD and the BASC

Levels of agreement between each clinical interview and each behavioral rating were determined by calculating product-moment correlations between symptom counts (i.e., the number of ODD symptoms endorsed on the ADIS–P and the DISC–IV and the number endorsed on the DBD), parent-endorsed interference and impairment ratings, clinical severity ratings, and BASC scale T-scores. Product-moment correlations between the interviews and the questionnaires varied among reporters (see Tables 4, 5, and 6). It is noteworthy that, for maternal-report, all of the correlations between the ADIS–P and the questionnaires were significant. All of the correlations between the DISC–IV and the questionnaires were also significant, except for the correlation between DISC–IV symptom count and DISC–IV impairment total.

As shown in Tables 4 and 5, the product-moment correlations between the ADIS–P and the other measures were often higher than those between the DISC–IV and the other variables. To test whether these differences were significant, dependent sample *t*-tests were conducted (see Table 7). Results indicated that although some differences were evident, they were not statistically significant, suggesting that the ADIS–P and the DISC–IV possess comparable relationships with the questionnaires, parent interference ratings, and clinical severity ratings included in this study.

Further analyses indicated that mothers, fathers, and teachers had variable degrees of agreement, dependent on the measures used (see Table 8). On the DBD, ratings among the reporters were not correlated. However, on the BASC, maternal ratings on the Aggression scale were related to paternal ratings on the same scale ($r = .64, p < .001$), as were maternal ratings on the Conduct Problems scale with paternal ratings ($r = .58, p < .001$) on the same scale. As shown in Table 8, ratings were correlated across measures within each category of reporter, suggesting consistency among the measures.

Finally, the product-moment correlations between interview and questionnaire data were typically higher for mothers than for fathers (see Table 7). To test whether these differences were significant, dependent sample *t*-tests were conducted on correlations between the questionnaire scales (i.e., the BASC Aggression and Conduct Problems scales and the DBD ODD subscale) and interview data (i.e., ADIS-P ODD symptoms, DISC-IV ODD symptoms, ADIS-P interference ratings, DISC impairment sums, ADIS-P clinician severity ratings, and consensus severity ratings). The correlation between the number of ODD symptoms endorsed on the DBD and the number of symptoms endorsed on the ADIS-P was significantly higher for mothers than for fathers, $t(43) = 2.60, p < .05$, as were the correlations between ADIS-P symptom totals and BASC Aggression T-scores, $t(43) = 2.12, p < .05$, and ADIS-P clinician severity ratings and DBD symptom totals, $t(43) = 2.14, p < .05$. However, the correlation between DISC-IV impairment totals and BASC Conduct Problems T-scores was higher for fathers than mothers, $t(43) = 2.08, p < .05$. All other correlations were not significantly different between mothers and fathers.

Discussion

The current study sought to examine the validity of the ADIS-P ODD module. It was hypothesized that the ADIS-P and the DISC-IV would demonstrate good levels of agreement at the symptom level and at the diagnostic level. Moreover, it was hypothesized that the number of symptoms endorsed on the ADIS-P ODD module would be correlated with the number of symptoms endorsed on the DISC-IV ODD module, the DBD ODD subscale, and the BASC Aggression and Conduct Problems scales. Finally, it was hypothesized that the parent-report interference rating on the ADIS-P would be correlated with the sum of parent-report DISC-IV impairment ratings.

Overall, results from this study indicate that the ODD module of the ADIS-P is comparable to the ODD module of the DISC-IV, with no order effects observed (see Table 1). Symptom-level agreement rates and kappa coefficients ranged from slight to moderate (see Table 2); however, the diagnosis-level agreement rates and kappa coefficients were, for the most part, in the moderate range (see Table 3), following guidelines proposed by Cicchetti and Feinstein (1990). As hypothesized, the highest diagnostic-level kappa coefficient and overall agreement rate resulted from determining DISC-IV diagnoses utilizing Impairment Algorithm D. This suggests that the impairment items on the DISC-IV are diagnostically relevant for ODD, although it remains unclear about which impairment algorithm is the best to use.

The correlation between symptoms endorsed on the ADIS-P and those endorsed on the DISC-IV was large; however, this correlation is lower than might have been expected inasmuch as the same symptoms were examined on both instruments and during the same assessment session. This finding suggests that although the symptom counts on the ADIS-P and the DISC-IV are related, they differ and are not interchangeable. Still, diagnostic rates were similar across the two instruments. The ADIS-P symptom count, interference rating, clinician severity rating, and consensus severity rating were correlated with the DISC-IV impairment total, the DBD ODD subscale, and the BASC parent-report Aggression and Conduct Problems scales (see Tables 4 and 5). Furthermore, correlations between the ADIS-P and the other measures did not differ from the correlations between the DISC-IV and the same measures (see Table 7), suggesting that the ODD modules of the ADIS and the DISC are measuring a similar construct (i.e., behavioral problems symptomatic of ODD).

There are several possible explanations for the observed differences among the correlations between the ADIS-P and the other measures and those between the DISC-IV and the same measures. As noted earlier, the DISC-IV alters the phrasing of some DSM symptoms, and the presentations of three of the eight *DSM-IV* symptoms of ODD are subdivided: the symptoms of defiance, anger, and spite. Furthermore, parents are given different directions for the two different interviews. For each behavior to count as a symptom on the DISC-IV, parents are asked whether the behavior has been occurring at a frequency of at least once a week for a period of at least six months; however, the ADIS-P counts the behavior as a symptom if the parent endorses it as occurring “a lot more than would be appropriate for his or her age.” These differences between the interviews might explain why symptom-level agreement rates and kappa

coefficients were variable among the eight ODD symptoms. In particular, these differences might have had an impact on the symptom of anger, which is presented differently by the ADIS-P than the DISC-IV. This symptom is split into two items on the DISC-IV: “Has he or she seemed mad at people or things?” and “Has he or she gotten angry because he or she thought things were unfair?” On the ADIS-P, this symptom is presented as, “Does he or she often seem as though he or she is angry at and resentful of other people?” The low agreement rates for this symptom suggest that these two DISC-IV items (mad at people or things) do not measure the symptom of anger in the same manner as does the ADIS-P item (mad at people only).

Results from the present study indicate that the various DISC-IV impairment algorithms lead to different agreement rates with the ADIS-P at the diagnostic level (see Table 3). Although differences in kappa coefficients were modest, Impairment Algorithm D appeared to have an advantage (i.e., have superior agreement with the ADIS-P ODD module) compared to the other impairment algorithms. It should be mentioned that Impairment Algorithm C, which requires at least one impairment item to be endorsed at a severe level, had the lowest kappa coefficient and negative and total agreement rates with the ADIS-P. Thus, these results indicate that the ADIS-P ODD module has the best agreement with the DISC-IV ODD module when Impairment Algorithm D is utilized.

That teacher-report data were not correlated with the other measures (see Tables 6 and 8) deserves some discussion. It is possible that the lack of corroboration between parents and teachers is due in part to some of the children in the sample exhibiting disruptive behaviors in the home but not in the classroom, although most children did evidence these symptoms across both settings. It is also possible that parents were more likely to endorse a symptom than were teachers (i.e., perhaps parents were stating that a behavioral problem had reached the threshold of being a clinical symptom when the teacher was reporting that the behavior was normative). It should be mentioned that the small sample size of teacher-report data decreased the power of the correlation analyses.

That three correlations between interview and questionnaire data (i.e., between the number of symptoms endorsed on the DBD and the number endorsed on the ADIS-P, between ADIS-P symptom totals and BASC Aggression T-scores, and between ADIS-P clinician severity ratings and DBD symptom totals) were higher for mothers than for fathers might be partly explained by the fact that mothers or female guardians participated in the interviews for all

of the cases, whereas fathers or male guardians participated in a minority of cases. Thus, the differences in correlations across parental report might have occurred because maternal report had more direct influence on interview data than paternal report had. These differences might also be due to the lower number of fathers who participated in the interviews. However, the correlation between DISC-IV impairment totals and BASC Conduct Problems T-scores was higher for fathers than for mothers, which suggests that this effect was not consistent across the measures.

Limitations

The current study's limitations include utilizing a small, clinical sample of children referred for treatment for ODD. Because parents were participating in the assessment as part of seeking treatment for their child's disruptive behavior, it is possible that they were more likely to view the child's behavior as pathognomic than they would be in other settings; consequently, they might have been more likely to endorse symptoms on the ODD modules of the ADIS-P and the DISC-IV. Because the majority of the children in the sample met criteria for a diagnosis of ODD on both the ADIS-P and the DISC-IV, it was not possible to conduct certain analyses (e.g., chi square), and kappa coefficients were attenuated due to skewed data (Cicchetti and Feinstein, 1990). The children who did not meet full criteria for ODD based on the ADIS-P and the DISC-IV were included in the study based on consensus diagnostic decisions. In consensus meetings, clinicians presented assessment data, including diagnostic impressions based on administration of the ADIS-P and the DISC-IV. Final diagnoses were determined under the supervision of a licensed clinical psychologist. The five cases that did not meet full criteria for ODD on both the ADIS-P and the DISC-IV (using Impairment Algorithm D) were assigned ODD diagnoses in the consensus meetings, based on additional assessment data and clinical judgment. Another limitation is that, although reliability was calculated for the ODD module of the ADIS-P, reliability statistics for DISC-IV diagnoses of ODD were not available for the present study. It should also be noted that the majority of the sample was Caucasian, which limits the generalizability of the findings to other ethnic samples. Another limitation is that many teachers did not complete the questionnaires, decreasing the sample size and the power of the analyses involving teacher-report variables. Finally, the ADIS-P and the DISC-IV were administered during the same assessment session, which might have inflated agreement rates. It

would be informative to examine the agreement rates between the ADIS–P and the DISC–IV after administering the modules one or two weeks apart.

Clinical Implications

The findings of this study have clinical implications for the use of the ADIS–P in the diagnostic assessment of ODD. The results presented here suggest that the ADIS–P ODD module is an acceptable instrument for diagnosing ODD, based on its comparability with the DISC–IV ODD module and its convergence with questionnaire and rating measures. Previous research has indicated that the ADIS–C/P is an appropriate diagnostic instrument for diagnosing anxiety disorders (Grills-Taquechel & Ollendick, 2008; Silverman et al., 2008; Wood et al., 2002), and one study (Jarrett et al., 2007) has suggested that the ADIS–C/P is valid for the assessment of Attention-Deficit/Hyperactivity Disorder. Taken together, the findings of the current study indicate that the ADIS–C/P could be particularly helpful for clinicians and researchers working with populations that exhibit ODD and comorbid anxiety disorders and Attention-Deficit/Hyperactivity Disorder.

References

- Achenbach, T. M. (1998). Diagnosis, assessment, taxonomy, and case formulations. In T. H. Ollendick & M. Hersen (Eds.) *Handbook of child psychopathology* (3rd ed.; pp. 63–87). New York: Springer.
- American Psychiatric Association (2000). *Diagnostic and statistical manual of mental disorders* (text revision). Washington, DC: Author.
- Barkley, R.A. (1997). *Defiant children: A clinician's manual for parent training* (2nd ed.). New York: Guilford.
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology*, *51*, 1173–1182.
- Chase, R. M., & Eyberg, S. M. (2008). Clinical presentation and treatment outcome for children with comorbid externalizing and internalizing symptoms. *Anxiety Disorders*, *22*, 273–282.
- Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of Clinical Epidemiology*, *43*, 551–558.
- Costin, J., & Chambers, S. M. (2007). Parent Management Training as a treatment for children with Oppositional Defiant Disorder referred to a mental health clinic. *Clinical Child Psychology and Psychiatry*, *12*, 511–524.
- Dewey, D., Kaplan, B. J., Crawford, S. G., & Fisher, G. C. (2001). Predictive accuracy of the Wide Range Assessment of Memory and Learning in children with Attention Deficit Hyperactivity Disorder and reading difficulties. *Developmental Neuropsychology*, *19*, 173–189.
- Frick, P. J., & McMahon, R. J. (2008). Child and adolescent conduct problems. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 41–66). New York: Oxford University Press.
- Friman, P. C., Handwerk, M. L., Smith, G. L., Larzelere, R. E., Lucas, C. P., & Shaffer, D. M. (2000). External validity of Conduct and Oppositional Defiant Disorders determined by the NIMH Diagnostic Interview Schedule for Children. *Journal of Abnormal Child Psychology*, *28*, 277–286.

- Geller, D. A., Biederman, J., Faraone, S., Spencer, T., Doyle, R., Mullin, B., et al. (2004). Re-examining comorbidity of Obsessive Compulsive and Attention-Deficit Hyperactivity Disorder using an empirically derived taxonomy. *European Child & Adolescent Psychiatry, 13*, 83–91.
- Grills, A. E., & Ollendick, T. H. (2003). Multiple informant agreement and the Anxiety Disorders Interview Schedule for Parents and Children. *Journal of the American Academy of Child & Adolescent Psychiatry, 42*, 30–40.
- Grills, A. E., & Ollendick, T. H. (2002). Issues in parent–child agreement: The case of structured diagnostic interviews. *Clinical Child and Family Psychology Review, 5*, 57–83.
- Grills-Taquechel, A., & Ollendick, T. H. (2008). Diagnostic interviewing. In M. Hersen & A. M. Gross (Eds.), *Handbook of clinical psychology* (Vol. 2, pp. 458–479). Hoboken, NJ: John Wiley & Sons, Inc.
- Hayes, M. A., McReynolds, L. S., & Wasserman, G. A. (2005). Paper and voice MAYSI-2: Format comparability and concordance with the Voice DISC-IV-IV. *Assessment, 12*, 395–403.
- Jarrett, M. A., Wolff, J. C., & Ollendick, T. H. (2007). Concurrent validity and informant agreement of the ADHD module of the Anxiety Disorders Interview Schedule for DSM-IV. *Journal of Psychopathology and Behavioral Assessment, 29*, 159–168.
- Jensen, P. S., Rubio-Stipec, M., Canino, G., Bird, H. R., Dulcan, M. K., Schwab-Stone, M. E., et al. (1999). Parent and child contributions to diagnosis of mental disorder: Are both informants always necessary? *Journal of the American Academy of Child & Adolescent Psychiatry, 38*, 1569–1579.
- Kamphaus, R. W., & Frick, P. J. (2005). *Clinical assessment of child and adolescent personality and behavior* (2nd ed.). New York: Springer.
- Landis, R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*, 159–174.
- Lemery-Chalfant, K., Schreiber, J. E., Schmidt, N. L., Van Hulle, C. A., Essex, M. J., & Goldsmith, H. H. (2007). Assessing internalizing, externalizing, and attention problems in young children: Validation of the MacArthur HBQ. *Journal of the American Academy of Child & Adolescent Psychiatry, 46*, 1315–1323.

- Lyneham, H. J., Abbott, M. J., & Rapee, R. M. (2007). Interrater reliability of the Anxiety Disorders Interview Schedule for *DSM-IV*: Child and Parent Version. *Journal of the American Academy of Child & Adolescent Psychiatry*, *46*, 731–736.
- McMahon, R. J., & Frick, P. J. (2005). Evidence-based assessment of conduct problems in children and adolescents. *Journal of Clinical Child and Adolescent Psychology*, *34*, 477–505.
- Merrell, K. W., Blade, R. L., Lund, J., & Kempf, K. K. G. (2003). Convergent and discriminant construct validity of the Internalizing Symptoms Scale for Children with the BASC–SRP–C. *Psychology in the Schools*, *40*, 139–144.
- Merydith, S. P. (2001). Temporal stability and convergent validity of the Behavior Assessment System for Children. *Journal of School Psychology*, *39*, 253–265.
- MTA Cooperative Group (2000). Anxiety as a predictor and outcome variable in the Multimodal Treatment Study of Children with ADHD (MTA). *Journal of Abnormal Child Psychology*, *28*, 527–541.
- Nemeroff, R., Levitt, J. M., Faul, L., Wonpat-Borja, A., Bufferd, S., Setterberg, S., et al. (2008). Establishing ongoing, early identification programs for mental health problems in our schools: A feasibility study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *47*, 328–338.
- Ollendick, T. H. (2007). Mediators, moderators, and treatment efficacy of two forms of psychosocial treatment for oppositional youth. National Institute of Mental Health Grant #1RO1 MH076141.
- Pelham, W. E., Gnagy, E. M., Greenslade, K. E., & Milich, R. (1992). Teacher ratings of DSM–III–R symptoms for the disruptive behavior disorders. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31*, 210–218.
- Reynolds, C. R., & Kamphaus, R. W. (1992). *Behavior Assessment System for Children–Second Edition*. Circle Pines, MN: American Guidance Service Publishing.
- Roberts, R. E., Roberts, C. R., & Xing, Y. (2007). Rates of DSM–IV psychiatric disorders among adolescents in a large metropolitan area. *Journal of Psychiatric Research*, *41*, 959–967.

- Schniering, C. A., Hudson, J. L., & Rapee, R. M. (2000). Issues in the diagnosis and assessment of anxiety disorders in children and adolescents. *Clinical Psychology Review, 20*, 453–478.
- Shaffer, D., Fisher, P., Lucas, C. P., Dulcan, M. K., & 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*, 28–39.
- Silverman, W. K., & Albano, A. M. (1996). *The Anxiety Disorders Interview Schedule for Children for DSM–IV: (Child and Parent Versions)*. San Antonio, TX: Psychological Corporation.
- Silverman, W. K., & Hinshaw, S. P. (2008). The second special issue on evidence-based psychosocial treatments for children and adolescents: A 10-year update. *Journal of Clinical Child & Adolescent Psychology, 37*, 1–7.
- Silverman, W. K., & Ollendick, T. H. (2005). Evidence-based assessment of anxiety and its disorders in children and adolescents. *Journal of Clinical Child and Adolescent Psychology, 34*, 380–411.
- Silverman, W. K., & Ollendick, T. H. (2008). Child and adolescent anxiety disorders. In J. Hunsley and E. J. Mash (Eds.), *A guide to assessments that work* (pp. 181–206). New York, NY: Oxford University Press.
- Silverman, W. K., Pina, A. A., & Viswesvaran, C. (2008). Evidence-based psychosocial treatments for phobic and anxiety disorders in children and adolescents. *Journal of Clinical Child & Adolescent Psychology, 37*, 105–130.
- Silverman, W. K., Saavedra, L. M., & Pina, A. A. (2001). Test-retest reliability of anxiety symptoms and diagnoses with the Anxiety Disorders Interview Schedule for *DSM–IV: Child and Parent Versions*. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*, 937–944.
- Wood, J. J., Piacentini, J. C., Bergman, R. L., McCracken, J., & Barrios, V. (2002). Concurrent validity of the anxiety disorders section of the Anxiety Disorders Interview Schedule for *DSM–IV: Child and Parent Versions*. *Journal of Clinical Child and Adolescent Psychology, 31*, 335–342.

Wright, K. D., Waschbusch, D. A., & Frankland, B. W. (2007). Combining data from parent ratings and parent interview when assessing ADHD. *Journal of Psychopathology and Behavioral Assessment*, 29, 141–148.

Table 1

Descriptive Statistics and t-Tests

Variables	Total Sample Means (SD)	ADIS–P First (<i>n</i> = 30)	DISC–IV First (<i>n</i> = 23)	<i>t</i> - values
1. ADIS–P Symptom Total	6.02 (1.51)	6.27 (1.44)	5.70 (1.58)	1.37
2. DISC–IV Symptom Total	6.28 (1.51)	6.33 (1.54)	6.22 (1.51)	.27
3. ADIS–P Parent Interference Rating	6.40 (1.35)	6.43 (1.36)	6.35 (1.37)	.23
4. ADIS–P Clinician Severity Rating	5.96 (1.21)	6.03 (1.10)	5.87 (1.36)	.49
5. ADIS–P Consensus Severity Rating	5.87 (1.04)	5.80 (1.00)	5.96 (1.11)	-.54
6. DISC–IV Impairment Total	9.06 (3.43)	8.93 (3.75)	9.22 (3.04)	-.30
7. Maternal DBD ODD Total	5.63 (1.78)	5.57 (1.87)	5.73 (1.70)	-.32
8. Paternal DBD ODD Total	4.17 (2.52)	4.22 (2.55)	4.11 (2.54)	.15
9. Teacher DBD ODD Total	1.53 (2.31)	1.16 (1.74)	2.08 (2.96)	-1.12
10. Maternal BASC Agg. T-score	70.75 (10.87)	70.80 (8.87)	70.70 (13.26)	.03
11. Paternal BASC Agg. T-score	65.25 (11.12)	64.43 (9.18)	66.40 (13.55)	-.60
12. Teacher BASC Agg. T-score	55.76 (12.02)	54.79 (11.91)	57.07 (12.50)	-.53
13. Maternal BASC CP T-score	70.26 (10.39)	69.57 (9.62)	71.17 (11.48)	-.55
14. Paternal BASC CP T-score	65.92 (13.64)	63.50 (12.79)	69.30 (14.39)	-1.47
15. Teacher BASC CP T-score	54.97 (10.22)	54.47 (10.06)	55.64 (10.76)	-.32

Note. DBD ODD totals were calculated by summing the number of symptoms endorsed at a 2 or 3 on a 0–3 rating scale, range 0–8. Agg. = Aggression Scale; CP = Conduct Problems Scale. All *p*-values $\geq .15$.

Table 2

Rates of Agreement and Kappa Coefficients for ADIS-P and DISC-IV Symptoms (N = 53)

Symptom	Positive Agreement	Negative Agreement	Total Agreement	κ
1. Temper	100%	17%	91%	.26**
2. Argue	96%	0%	96%	—
3. Defiance	94%	75%	93%	.56***
4. Annoy	93%	63%	79%	.57***
5. Blame	97%	47%	81%	.51***
6. Grouchy	78%	86%	79%	.41**
7. Anger	55%	67%	57%	.09
8. Spite	69%	78%	74%	.47**

Note. The DISC-IV was the criterion measure. The kappa coefficient could not be calculated for the Argue symptom, because the item was endorsed positively on the DISC-IV in every case.

** $p < .01$. *** $p < .001$.

Table 3

*Rates of Agreement and Kappa Coefficients for ADIS–P and DISC–IV Diagnostic Decisions**(N = 53)*

DISC–IV Impairment Algorithm	Positive Agreement	Negative Agreement	Total Agreement	κ
Impairment Algorithm A	96%	67%	94%	.54***
Impairment Algorithm B	98%	60%	94%	.64***
Impairment Algorithm C	100%	36%	87%	.48***
Impairment Algorithm D	98%	75%	96%	.73***
Impairment Excluded	96%	67%	94%	.54***

Note. The DISC–IV was the criterion measure. To assign a diagnosis, Impairment Algorithm A requires at least one of the impairment items to be endorsed at a moderate or severe level (2 or 3 on 0–3 scale). Impairment Algorithm B requires at least two of the impairment items to be endorsed at a moderate or severe level. Impairment Algorithm C requires at least one impairment item to be endorsed at a severe level. Impairment Algorithm D will assign a diagnosis if criteria for either Impairment Algorithms B or C are met. Diagnoses assigned without including any impairment criterion are based solely on the number of symptoms that the parent endorsed as occurring for 6 months or longer.

*** $p < .001$.

Table 4

Correlations among Maternal-Report Questionnaire Scales, Interference Ratings, Interview Symptom Counts, and Clinician Severity Ratings (n = 52)

Measure	1	2	3	4	5	6	7	8	9
1. ADIS-P Symptoms	—	.64***	.72***	.55***	.41**	.45**	.32*	.77***	.58***
2. DISC-IV Symptoms		—	.56***	.41**	.33*	.31*	.21	.66***	.60***
3. DBD ODD Total			—	.57***	.53***	.35*	.22	.62***	.53***
4. Aggression				—	.52**	.26	.33*	.46**	.33*
5. Conduct Problems					—	.27	.37**	.41**	.40**
6. ADIS-P Interference						—	.48***	.52***	.56***
7. DISC-IV Impairment							—	.47***	.48***
8. ADIS-P CSR								—	.81***
9. Consensus Severity Rating									—

Note. Interview symptom counts were used in correlations. Aggression and Conduct Problems scales are from the BASC. DBD ODD totals were calculated by summing the number of symptoms endorsed at a 2 or 3 on a 0–3 rating scale, range 0–8.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 5

Correlations among Paternal-Report Questionnaire Scales, Interference Ratings, Interview Symptom Counts, and Clinician Severity Ratings (n = 46)

Measure	1	2	3	4	5	6	7	8	9
1. ADIS-P Symptoms	—	.57***	.32*	.37*	.25	.39**	.35*	.74***	.51***
2. DISC-IV Symptoms		—	.27	.24	.30*	.24	.23	.63***	.54***
3. DBD ODD Total			—	.58***	.61***	.32*	.43**	.23	.37*
4. Aggression				—	.61***	.20	.50***	.40**	.50***
5. Conduct Problems					—	.33*	.61***	.34*	.47**
6. ADIS-P Interference						—	.46**	.46**	.50***
7. DISC-IV Impairment							—	.48**	.50***
8. ADIS-P CSR								—	.79***
9. Consensus Severity Rating									—

Note. Interview symptom counts were used in correlations. Aggression and Conduct Problems scales are from the BASC. DBD ODD totals were calculated by summing the number of symptoms endorsed at a 2 or 3 on a 0–3 rating scale, range 0–8.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 6

Correlations among Teacher-Report Questionnaire Scales, Interference Ratings, Interview Symptom Counts, and Clinician Severity Ratings (n = 31)

Measure	1	2	3	4	5	6	7	8	9
1. ADIS-P Symptoms	—	.62***	.03	-.10	.05	.48**	.48**	.76***	.60***
2. DISC-IV Symptoms		—	-.09	-.20	-.06	.24	.20	.69***	.59**
3. DBD ODD Total			—	.81***	.46**	.29	.25	.16	.13
4. Aggression				—	.65***	.17	.10	-.00	-.02
5. Conduct Problems					—	.29	.17	.11	.22
6. ADIS-P Interference						—	.63***	.50**	.49**
7. DISC-IV Impairment							—	.57**	.56**
8. ADIS-P CSR								—	.83***
9. Consensus Severity Rating									—

Note. Interview symptom counts were used in correlations. Aggression and Conduct Problems scales are from the BASC. DBD ODD totals were calculated by summing the number of symptoms endorsed at a 2 or 3 on a 0–3 rating scale, range 0–8.

** $p < .01$. *** $p < .001$.

Table 7

t-Tests for ADIS-P and DISC-IV Correlations

Variables	<i>df</i>	Correlation with ADIS-P	Correlation with DISC-IV	<i>t</i> -values	<i>p</i> -values
Maternal DBD ODD Total	49	.72	.56	1.89	.06
Paternal DBD ODD Total	43	.32	.27	0.31	.75
Maternal Aggression	50	.56	.42	1.39	.17
Paternal Aggression	43	.37	.24	1.01	.32
Maternal Conduct Problems	50	.42	.34	0.76	.45
Paternal Conduct Problems	43	.25	.30	-0.35	.73
ADIS Interference	50	.43	.29	1.26	.22
DISC Impairment Total	50	.32	.21	0.90	.37
Clinician Severity Rating	50	.77	.66	1.46	.15
Consensus Severity Rating	50	.58	.59	-0.12	.91

Note. Interview symptom counts were used in correlations. Aggression and Conduct Problems scales are from the BASC. DBD ODD totals were calculated by summing the number of symptoms endorsed at a 2 or 3 on a 0–3 rating scale, range 0–8.

Table 8

Correlations among Adult Reporters on Questionnaire Scales

Measure	1	2	3	4	5	6	7	8	9
1. Maternal DBD	—	.21	-.11	.57***	.33*	-.14	.53***	.27	-.06
2. Paternal DBD		—	-.29	.31*	.58***	-.23	.45**	.61***	.01
3. Teacher DBD			—	-.30	-.18	.81***	-.21	-.04	.46**
4. Maternal Agg.				—	.64***	-.14	.52***	.33*	.20
5. Paternal Agg.					—	-.03	.38**	.60***	.16
6. Teacher Agg.						—	-.25	-.07	.66***
7. Maternal CP							—	.58***	.13
8. Paternal CP								—	.18
9. Teacher CP									—

Note. CP = Conduct Problems scale. The BASC Aggression and Conduct Problems scale scores were converted into T-scores. DBD Total is the symptom count from the ODD items from the DBD. Values for n ranged from 29–53 due to missing or incomplete data.

* $p < .05$. ** $p < .01$. *** $p < .001$.

Appendix—Revised DBD

This appendix contains the revised DBD items that pertain to ODD.

Instructions: Circle the number that best describes your child's behavior at home over the past week.

	Never or Rarely	Sometimes	Often	Very Often
1. Loses temper	0	1	2	3
2. Argues with adults	0	1	2	3
3. Actively defies or refuses to comply with adults' requests or rules	0	1	2	3
4. Deliberately annoys people	0	1	2	3
5. Blames others for his/her mistakes or misbehavior	0	1	2	3
6. Is touchy or easily annoyed by others	0	1	2	3
7. Is angry and resentful	0	1	2	3
8. Is spiteful or vindictive	0	1	2	3