Journal of Psychopathology and Behavioral Assessment, Vol. 27, No. 3, September 2005 (© 2005)
DOI: 10.1007/s10862-005-0632-8

Parent–Adolescent Agreement on Disruptive Behavior Symptoms: A Multitrait-Multimethod Model

Cynthia M. Hartung,[1,4] Denis M. McCarthy,[2] Richard Milich,[3] and Catherine A. Martin[3]

Accepted June 22, 2004

This study examined parent–adolescent agreement on reports of inattention, hyperactivity/impulsivity, and conduct disorder in 203 adolescents (94 girls, 109 boys) ranging in age from 13 to 18 years ($M = 15.21$, $SD = 1.37$). Results of confirmatory factor analyses provided additional evidence of construct validity for these traits in adolescents. Internal consistency was examined for parent reports and adolescent self-reports. In addition, correlational analyses were used in a multitrait-multimethod format (MTMM; D. T. Campbell & D. W. Fiske, 1959) to examine convergent and discriminant validity. Results showed that parents provided more consistent and valid reports of inattention and hyperactivity/impulsivity, whereas adolescents provided more consistent and valid reports of conduct disorder. In addition, interviews produced higher levels of convergence than rating scales. These findings are discussed in terms of implications for assessment of disruptive behavior disorders.

KEY WORDS: disruptive behavior disorders; attention-deficit/hyperactivity disorder (ADHD); conduct disorder (CD); adolescents; parent–adolescent agreement.

Diagnostic information about children and adolescents is typically collected from multiple informants (e.g., parents, teachers, and self-reports). One of the challenges confronting researchers in child and adolescent psychopathology is the modest interrater agreement between different informants (e.g., Achenbach, McConaughy, & Howell, 1987). Considerable research has been conducted on interrater reliability and validity of reports of child and adolescent symptomatology (e.g., Achenbach et al., 1987; Lahey et al., 2000; Tildesley, Hops, Ary, & Andrews, 1995). There is general agreement that diagnoses should be based on multiple informants (Bird, Gould, & Staghezza, 1992; Hart, Lahey, Loeber, & Hanson, 1994; Lahey et al., 1996) including mothers, fathers, teachers, and even psychological examiners (Willcutt, Hartung, Lahey, Loney, & Pelham, 1999). However, several questions remain concerning the reliability and validity of reports when assessing disruptive behavior disorder symptoms. In particular, reliability and validity may vary by informant (e.g., parent, teacher, self-report), format (e.g., interviews vs. rating scales), and/or behavior assessed (e.g., hyperactivity vs. conduct problems).

To evaluate reports of child and adolescent symptoms, Achenbach et al. (1987) conducted a meta-analysis of agreement between multiple-informants. Studies of clinic-referred and nonreferred youth between the ages of 2 and 19 were included. Modest agreement between two parents' reports of their child's symptomatology (mean $r = .59$) was found. In contrast, the correlation between one parent and the child or adolescent was lower (mean $r = .25$). Thus, across a variety of studies, including participants of differing referral status and age, child and adolescent informants appeared to be less reliable than parents as evidenced by lower interrater agreement when self-reports were included.

However, several possible explanations have been offered for the apparently lower reliability of child and adolescent self-reports. The most obvious explanation is

[1]Oklahoma State University, Stillwater, Oklahoma.
[2]University of Missouri-Columbia, Columbia, Missouri.
[3]University of Kentucky, Lexington, Kentucky.
[4]To whom correspondence should be addressed at Department of Psychology, Oklahoma State University, 215 North Murray, Stillwater, Oklahoma 74078; e-mail: cmhart@okstate.edu.

0882-2689/05/0900-0159/0 © 2005 Springer Science+Business Media, Inc.

that children, especially younger ones, are inaccurate, unreliable informants (Edelbrock, Costello, Dulcan, Conover, & Kala, 1986) whereas adolescents may be as reliable as adults (e.g., Tildesley et al., 1995). A second possible explanation for the apparently lower reliability of child and adolescent self-reports, when compared to parents and teachers, is that all informants are providing valid, but different, information (Achenbach et al., 1987; Greenbaum, Dedrick, Prange, & Friedman, 1994). As Greenbaum et al. (1994) describe it, cross-situational specificity may account for the modest interrater agreement. That is, children may behave differently in different situations (e.g., home, school, with peers) and informants are rating behavior on the basis of observations in different settings. Specifically, teachers are describing the child's behavior in the classroom, parents are describing the child's behavior at home, and children and adolescents are describing their own behavior across school, home, and unstructured peer interactions. If this explanation is accurate, then interrater agreement should not be considered a measure of reliability. Lahey et al. (2000) found poor agreement between youth (ages 9–17) and their parents on ratings of conduct problems. Further analyses suggested that youth, especially older ones, reported higher levels of conduct problems than parents. It is possible that parents are accurately reporting on the basis of youth's behaviors at home, whereas youth may be accurately reporting on the basis of their behaviors across multiple settings. Thus, both parents and youths may be reporting accurately but youths may be privy to more information about conduct problems than parents.

A third explanation for the apparently lower reliability of self-reports is that reliability may vary by informant (e.g., child, parent) and/or disruptive behavior disorder (e.g., conduct disorder [CD] vs. attention-deficit/hyperactivity disorder [ADHD]). That is, self-reports may be less reliable for some disorders but not all. Andrews, Garrison, Jackson, Addy, and McKeown (1993) found low agreement for CD. In contrast, Cantwell, Lewinsohn, Rohde, and Seeley (1997) found good parent–adolescent agreement for ADHD and excellent agreement for CD. Similarly, Thurber and Snow (1990) found good agreement between mother–daughter reports of CD but not for aggression. Loeber, Green, Lahey, and Stouthamer-Loeber (1989) concluded that parents were better informants of hyperactivity, inattentiveness, and oppositional behavior than were boys. Nonetheless, both boys and parents were reasonably accurate in reporting CD symptoms. In another study of adolescents, Crowley, Mikulich, Ehlers, Whitmore, and MacDonald (2001) found that self-reports discriminated patients from controls on the basis of CD and substance use disorders but not on the basis of ADHD. These findings illustrate the importance of continuing to study data from multiple informants by focusing on specific disorders and symptomatology.

Another potentially important issue, which has received limited attention at best, is that the reliability of reports may be impacted by the format (e.g., rating scales vs. interview) used to collect the data. Each type of assessment instrument is noted to have certain strengths and weaknesses. For example, rating scales are cost efficient and require little professional time, whereas interviews allow for more in-depth exploration of informant responses and the incorporation of clinical judgment. Literature searches did not identify any studies that have compared the reliability of interviews to rating scales. Nonetheless, knowledge of the comparative reliability and validity of these two data collection formats would allow researchers and clinicians to make informed decisions about the most efficient and accurate method of collecting diagnostic information.

In summary, research on the reliability of disruptive behavior disorder symptom reports in children and adolescents suggests that parents may be more reliable than adolescents who in turn may be more reliable than children. However, there is some evidence for differential agreement as a function of disorder assessed. For example, children and adolescents appear to be more reliable informants in reports of CD than reports of ADHD. The present study was designed to examine the correspondence of reports for three disruptive behavior traits (i.e., inattention [IA], hyperactivity/impulsivity [HI], and CD) using two informants (i.e., parents and adolescents) and two formats (i.e., rating scales and interviews). Two statistical approaches were also employed. First, confirmatory factor analysis (CFA) was used to examine the construct validity of the three traits. CFA approaches to multitrait-multimethod (MTMM) models allow for direct tests of the convergent validity, discriminant validity, and method effects of the full set of assessment measures (Byrne, 1994; Kenny & Kashy, 1992; Marsh & Grayson, 1995; Widaman, 1985). Second, internal consistency values and correlations were computed and presented in a traditional MTMM matrix (Campbell & Fiske, 1959). This approach allows for the examination of the reliability and validity evidence for each method/trait combination. For these analyses, a full model was specified to test for overall convergence and discrimination among traits, and the magnitude of method effects.

Specific hypotheses were that: (1) The construct validity for ADHD and CD would be strong, as evidenced by strong convergent and discriminant validity and weak method effects; (2) Parent reports of ADHD would be more reliable, as measured by internal consistency, and

more valid, as measured by convergent and discriminant validity, than adolescent reports of ADHD (Achenbach et al., 1987; Andrews et al., 1993; Crowley et al., 2001; Loeber et al., 1989); (3) Adolescent reports of CD would be more reliable and valid than parent reports of CD (Achenbach et al., 1987; Andrews et al., 1993; Crowley et al., 2001; Loeber et al., 1989); (4) Interviews would be more reliable than the rating scales regardless of the trait assessed.

METHOD

Participants in the study were 203 adolescents (ages 13–18; $M = 15.21$, $SD = 1.37$). Ninety-four were adolescent girls (46%) and 109 were adolescent boys (54%). In addition, 91 adolescents were clinic-referred (45%) with psychiatric or behavioral diagnoses; the remaining 112 adolescents were nonreferred (55%). The clinic-referred adolescents were recruited from: (1) outpatient psychiatry and psychology clinics, (2) a residential treatment center, and (3) a parent support group. The nonreferred adolescents were recruited through a newspaper announcement. Participants with estimated full scale IQ scores <80 were excluded from analyses ($n = 10$). Full scale IQ was estimated using short forms of either the Wechsler Intelligence Scale for Children—Third Edition (WISC-III; Wechsler, 1991) or the Wechsler Adult Intelligence Scale—Revised (WAIS-R; Wechsler, 1981) depending on age.

For each adolescent a parent or legal guardian also participated. Of the 203 parent participants, 168 were mothers (83%), 32 were fathers (16%), and 3 were grandparents (1%). As a measure of socioeconomic status, parents were asked to report their educational levels. On average, parents had received 3.5 years of college education ($M = 15.56$, $SD = 2.33$). With regard to ethnicity, parents reported that 186 adolescent participants were non-Hispanic White (92%) and that 15 were members of minority groups (7%); two families chose not to report ethnic background (1%).

After consent and assent were obtained from the parents and adolescents, respectively, adolescents and parents completed the first of two assessment procedures. (i.e., interview or rating scales). These assessment procedures were administered in a counterbalanced order. As part of a larger study, adolescents completed the first of two laboratory tasks (see Hartung, Milich, Lynam, & Martin, 2002). While these tasks were administered to the adolescent, the second assessment procedure was administered to the parent. After the first laboratory task was administered, adolescents completed the second assessment procedure followed by the second laboratory task. Following a break, brief tests of cognitive and reading abilities (as part of the larger study) were administered to the adolescent. To ensure that experimenters remained blind to diagnostic status, three experimenters worked with each family. One experimenter administered the experimental tasks and cognitive tests to the adolescent, a second experimenter interviewed the adolescent, and a third experimenter interviewed the parent. Parent and adolescent interviewers were graduate students in a clinical psychology doctoral program with clinical interviewing experience.

The study examined reports of ADHD and CD. Criteria for ADHD in the *Diagnostic and statistical manual of mental disorders* (*DSM-IV*; American Psychiatric Association, 1994), are divided into two categories (i.e., IA and HI). Therefore, in the current study, reports of IA, HI, and CD were examined.

Two formats were used with two informants to assess the presence of IA, HI, and CD resulting in four assessment methods: (1) adolescent interviews, (2) parent interviews, (3) adolescent ratings, and (4) parent ratings. All assessment methods included the *DSM-IV* criteria for ADHD and CD. It should be noted that these four methods were not completely independent because informant (i.e., parent and adolescent) was crossed with format (i.e., interview and rating scale).

The rating scale used in this study was the Disruptive Behavior Disorders Checklist (DBD; Pelham, Gnagy, Greenslade, & Milich, 1992). On this checklist each *DSM-IV* criterion is listed and informants were required to indicate whether each criterion was "not at all," "just a little," "pretty much," or "very much" true for the adolescent. "Pretty much" and "very much" responses were considered positive endorsements of a particular criterion (see Pelham et al., 1992). Pelham et al. used the rating scale format to collect normative data from teachers and obtained coefficient alphas of .96 and .75 for ADHD and CD diagnoses, respectively.

A semi-structured interview version of the DBD checklist was used. By adapting the DBD checklist for use as an interview, the similarity of the items across measures was maximized. This is important because one goal of this study was to compare the reliability and validity of these two assessment formats. Given that the items are identical, differences in reliability and validity can be attributed solely to the format. During the interview, informants were read each criterion and asked whether the adolescent exhibited the symptom. If the informant indicated the symptom was problematic, he/she was asked to give examples of the behavior. The *DSM-IV* description for ADHD, for example, requires that symptoms be "maladaptive and inconsistent with

developmental level." Interviewers then used this information to determine, on the basis of *DSM-IV* standards, whether the response indicated presence or absence of the symptom. To measure intercoder reliability, 10% of the interviews ($n = 20$) were recoded by another interviewer using audiotapes and Kappa coefficients were calculated. For adolescent interview, the average Kappa was .92 for IA, .88 for HI, and .86 for CD. For parent interview, the average Kappa was .92 for IA, .92 for HI, and .88 for CD.

In assessing the traits, the *DSM-IV* criteria for ADHD and CD were used in the interviews and rating scales. ADHD symptoms were grouped on the basis of *DSM-IV* IA and HI dimensions. The symptoms for each of the three traits were then averaged to form three composite variables (i.e., IA, HI, and CD). These composites were crossed with the four methods (i.e., parent interview, parent rating, adolescent interview, adolescent rating) to form 12 variables.

A series of four nested models were compared to conduct a MTMM analysis using CFA. The first model was the Freely Correlated Traits/Freely Correlated Methods Model (Model 1), or the full MTMM model. In this model, each of the 12 observed variables was an indicator of a trait latent variable, a method latent variable, and a unique component (see Fig. 1).

The second model was the Freely Correlated Methods/No Traits Model (Model 2). This model was a measurement model where each of the 12 observed variables was an indicator of only a method latent variable and a unique error term. Byrne (1994) recommended comparing the Freely Correlated Methods/No Traits Model (Model 2) to the Freely Correlated Traits/Freely Correlated Methods Model (Model 1) as a measure of convergent validity. She defined convergent validity as "the extent to which independent measures of the same trait are correlated" (p. 131). The chi-square difference test between Models 1 and 2 will reveal the extent to which the traits account for the variance in the model. That is, large chi-square difference values would suggest that the traits account for a great deal of variance and provide evidence for convergent validity.

The third model was the Perfectly Correlated Traits/Freely Correlated Methods Model (Model 3). This model had both trait and method latent variables specified, however, the trait factors were specified to be perfectly correlated (i.e., set to 1). As recommended by Byrne (1994) and Widaman (1985), Model 3 was compared to Model 1 to measure discriminant trait validity. Byrne defined discriminant trait validity as "the extent to which independent measures of different traits are correlated" (p. 132). A large chi-square difference value would provide evidence for discriminant trait validity.

Finally, a fourth model, the Perfectly Correlated Methods/Freely Correlated Traits Model (Model 4) was estimated. This model had both trait and method factors specified, with the method factors perfectly correlated. As recommended by Byrne (1994) and Widaman (1985), Model 4 was compared to Model 1 to look for possible method effects. Large chi-square difference values would suggest that method effects are not significant.

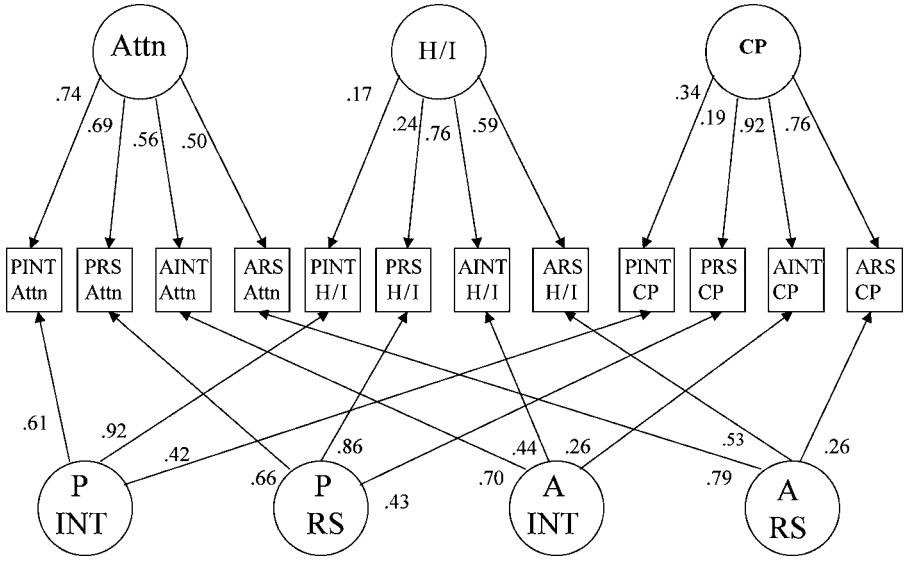

Fig. 1. Freely Correlated Traits/Freely Correlated Methods Model (Model 1). IA = inattention; H/I = hyperactivity/impulsivity; CD = conduct disorder.

Table I. Number and Percentage of Participants Meeting *DSM-IV* DBD Criteria by Method and Referral Status

	ADHD-only		CD-only		ADHD + CD		Total	
	n	(%)	*n*	(%)	*n*	(%)	*n*	(%)
Clinic-referred adolescents ($n = 91$)								
Adolescent interviews	8	(8.8)	11	(12.1)	4	(4.4)	23	(25.3)
Adolescent ratings	13	(14.3)	1	(1.1)	2	(2.2)	16	(17.6)
Parent interviews	45	(49.5)	4	(4.4)	8	(8.8)	57	(62.6)
Parent ratings	46	(50.5)	5	(5.5)	3	(3.3)	54	(59.3)
Non-referred adolescents ($n = 112$)								
Adolescent interviews	0	(0.0)	6	(5.4)	0	(0.0)	6	(5.4)
Adolescent ratings	2	(1.8)	2	(1.8)	0	(0.0)	4	(3.6)
Parent interviews	1	(0.9)	2	(1.8)	0	(0.0)	3	(2.7)
Parent ratings	3	(0.9)	0	(0.0)	0	(0.0)	3	(0.9)

Note. ADHD = Attention-deficit/hyperactivity disorder; CD = Conduct disorder; ADHD + CD = meets criteria for both ADHD and CD.

Analyses were performed to assess the correspondence between parent and adolescent reports of DBD symptomatology. A series of CFA analyses were conducted to assess construct validity (Byrne, 1994; Widaman, 1985). In addition, reliability of informants and formats was assessed by examining internal consistency. Finally, correlational analyses were conducted to assess convergent and discriminant validities. Participants were collapsed across referral status for all analyses.[5] Given that multiple statistical comparisons were conducted, $p < .001$ was established as the cutoff for significance.

RESULTS

The diagnostic description of the sample is shown in Table I. On the basis of adolescent interview, 25.3% of the clinic-referred and 5.4% of the nonreferred adolescents met diagnostic criteria for ADHD and/or CD. For adolescent rating, 17.6% of the clinic-referred and 3.6% of the nonreferred adolescents met diagnostic criteria. On the basis of parent interview, 62.6% of the clinic-referred and 2.7% of the nonreferred adolescents met diagnostic criteria. For parent rating, 59.3% of the clinic-referred and 0.9% of the nonreferred adolescents met diagnostic criteria. Although, parent reports resulted in more adolescents meeting criteria for ADHD than did adolescent self-reports ($\chi^2[1, N = 202] = 12.12, p < .001$), self-reports resulted in more adolescents meeting criteria for CD than did parent reports ($\chi^2[1, N = 202] = 17.22, p < .001$).

Mean levels of DBD symptoms for clinic-referred and nonreferred adolescents are presented in Table II. Paired samples *t* tests were conducted to examine differences in symptomatology reported by parents and adolescents using interviews and rating scales. First, symptom levels by informants were examined. For clinic-referred participants, parents reported significantly more IA and HI, but not more CD, symptoms than did adolescents. For nonreferred participants, adolescents reported more CD symptoms than did parents on the basis of the interviews but not rating scales. Second, symptom levels by formats were examined. For clinic-referred and nonreferred participants, the interview format resulted in significantly more CD symptoms, by both parents and adolescents, than did the rating scale format. Conversely, for clinic-referred and nonreferred participants, the rating scale format resulted in significantly more HI symptoms being reported by adolescents than did the interview format. For IA, clinic-referred adolescents reported significantly more symptoms via interview than rating scale whereas nonreferred adolescents reported significantly more symptoms via rating scale than

[5] Additional analyses were run to further justify collapsing across referral status. When CFA analyses were run using only the clinic-referred participants, all model comparisons were the same, with Model 1 having the best fit to the data. When the correlational analyses were run on the clinic-referred participants only, results were similar to those for the entire sample but sometimes did not reach the same level of significance due to lower power. However, most of the critical findings were maintained in this group ($ps < .001$). Specifically, for the clinic group, parent reports of CD were significantly less internally consistent than parent reports of IA ($z = 5.73$ and 6.12) and HI ($z = 3.52$ and 4.03) as measured by rating scales and interviews, respectively. In addition, adolescent reports of CD were significantly more internally consistent than parent reports as measured by rating scales ($z = 4.03$) and interviews ($z = 3.29$). Again, within-category correlations for both adolescents and parents were significant and parent reports resulted in lower correspondence for CD than for IA ($z = 5.11$) or HI ($z = 3.14$). In addition, adolescents demonstrated significantly higher correspondence for CD than did parents ($z = 3.64$).

Table II. Mean Number of Symptoms Endorsed and Comparisons of Reported Levels by Method

	M (*SD*)		
	IA	HI	CD
Clinic-referred adolescents ($n = 91$)			
Adolescent interviews	2.36 (2.52)$_a$	1.14 (1.59)$_a$	0.97 (1.79)$_a$
Adolescent ratings	1.95 (2.48)$_a$	1.60 (1.88)$_{a,b}$	0.51 (1.63)$_b$
Parent interviews	5.32 (3.29)$_b$	2.34 (2.45)$_b$	1.00 (1.25)$_{a,b}$
Parent ratings	4.99 (3.46)$_b$	2.55 (2.61)$_b$	0.65 (1.20)$_{a,b}$
Non-referred adolescents ($n = 112$)			
Adolescent interviews	0.61 (1.17)$_a$	0.47 (0.88)$_a$	0.39 (0.96)$_a$
Adolescent ratings	0.81 (1.41)$_a$	1.04 (1.17)$_b$	0.13 (0.49)$_b$
Parent interviews	0.77 (1.52)$_a$	0.38 (0.85)$_a$	0.21 (0.60)$_{a,b}$
Parent ratings	0.64 (1.50)$_a$	0.40 (0.93)$_a$	0.04 (0.23)$_b$

Note. IA = Inattention (9 symptoms possible); HI = hyperactivity/impulsivity (9 symptoms possible); CD = conduct disorder (15 symptoms possible). Means for the same trait without any common subscripts are significantly different based on two-tailed, paired samples *t*-tests ($p < .001$).

interview. Parent reports did not differ by format for HI or IA.

Analyses were conducted to test for significant interactions between four demographic variables (i.e., estimated full scale IQ, age, parent education, and sex) and the four methods (i.e., adolescent ratings, adolescent interviews, parent ratings, and parent interviews) on DBD symptom levels. The only demographic variable that significantly interacted with method to effect DBD symptoms levels was sex. For IA, parent ratings ($t = 3.29$, $p < 001$) and interviews ($t = 3.32$, $p < .001$) resulted in higher levels for adolescent boys than for girls. However, adolescent ratings and interviews did not result in significant sex differences on IA. For HI, only parent interviews resulted in higher levels for boys than for girls ($t = 3.28$, $p < .001$). For conduct problems, no significant sex by method differences were found. Although these findings are notable, the analyses in the current study could not be conducted separately by sex due to lack of power.

Confirmatory factor analysis was conducted using the maximum likelihood method of estimation, which has been found to be reasonably robust to violations of its assumptions, such as small sample size (Hu, Bentler, & Kano, 1992). For goodness of fit indexes, the chi-square (χ^2), Non-Normed Fit Index (NNFI), and the Comparative Fit Index (CFI) were used. As recommended by Hoyle and Panter (1995), one type-2 index (i.e., NNFI) and one type-3 index (i.e., CFI) were included. CFI or NNFI values >.90 indicate good fit between the data and the model (Byrne, 1994). A series of model comparisons were used to evaluate convergent and discriminant validity of traits, as well as size of method effects. Table III presents the chi-square values for the four CFA models, along with chi-square difference tests comparing models. Fit indices indicated that only Model 1 fit the data well.

The Freely Correlated Methods/No Traits Model (Model 2) was compared to the Freely Correlated Traits/Freely Correlated Methods Model (Model 1) as a measure of convergent validity. As shown in Table III, the chi-square difference test between Models 1 and 2 suggested that Model 1 was a very significant improvement over Model 2 ($\chi^2[15, N = 203] = 403.25$, $p < .001$), suggesting that the traits account for a significant amount of variance and providing evidence for convergent validity. Comparing Model 3 (Perfectly Correlated Traits/Freely Correlated Methods Model) to Model 1 is a measure of discriminant trait validity. As shown in Table III, the chi-square difference test between

Table III. Model Summary

Model	χ^2	*df*	CFI	NNFI
Model 1	70.61	33	.98	.95
Model 2	473.86	48	.72	.61
Model 3	780.28	36	.52	.10
Model 4	838.90	39	.47	.10
Test of convergent validity				
Model 1 vs. Model 2	403.25	15		
Tests of discriminant validity				
Model 1 vs. Model 3	709.67	3		
Model 1 vs. Model 4	768.29	4		

Note. All χ^2 values are significant ($p < .001$). Model 1 = Three trait factors, four method factors, three pairs of trait covariances estimated, six pairs of method covariances estimated. Model 2 = No trait factors, four method factors, six pairs of method covariances estimated. Model 3 = Perfectly correlated traits, freely correlated methods. Model 4 = Freely correlated traits, perfectly correlated methods.

Table IV. Correlations Between Factors for Model 1

	IA	HI	PI	PR	AI
HI	.45				
CD	$.16^{ns}$	.57			
PR			.96		
AI			.31	$.19^{ns}$	
AR			$.07^{ns}$	$-.02^{ns}$	.90

Note. All values are significant at $p < .001$ unless otherwise specified; ns = not significant. IA = inattention; HI = hyperactivity/impulsivity; CD = conduct disorder; PI = Parent interviews; PR = Parent ratings; AI = Adolescent interviews; AR = Adolescent ratings.

Models 1 and 3 suggested that Model 1 was a significant improvement over Model 3 ($\chi^2[3, N = 203] = 709.76$, $p < .001$) and provided evidence for discriminant trait validity.

Model 4 (Perfectly Correlated Methods/Freely Correlated Traits Model) was compared to Model 1 to look for possible method effects. Again, a χ^2 difference test between Models 1 and 4 suggested that Model 1 was a significant improvement over Model 4 ($\chi^2[4, N = 203] = 768.29$, $p < .001$) providing evidence against significant method effects. The correlations among trait and method factors in Model 1 were then examined as evidence of discriminant validity (see Table IV). Low factor correlations provide evidence for discriminant validity. HI was correlated with both IA and CD. The IA and CD correlation was not significant. Three of the six method factor correlations were significant.

The internal consistencies of the DBD dimensions are denoted in Table V in parentheses. Internal consistency was statistically compared between informants. In order to test the significance of the difference between coefficient alpha values, Fisher's z transformations were used to compare independent correlations (Cohen and Cohen, 1983). Parent reports of IA and HI were significantly (ps $< .001$) more internally consistent than adolescent reports of IA ($z = 5.50$ and 5.02) and HI ($z = 4.83$ and 4.28) as measured by ratings scales and interviews, respectively. However, for CD, adolescent reports were significantly more internally consistent than parent reports as measured by rating scales and interviews ($z = 4.80$ and 3.55). Internal consistency was also statistically compared between formats. There were no significant differences ($p < .001$) between formats for adolescents or parents.

Pearson correlational analyses are also reported in Table V in a MTMM format. Whereas the CFA analyses provided evidence of construct validity for the traits and the internal consistencies provided reliability data for informants and formats, the MTMM analyses allow a comparison of the relative validity of informants and formats for measuring DBD traits. Cross-category (e.g., parent interview-IA with parent rating-CD) and within-category (e.g., parent interview-IA with parent rating-IA) correlations were examined. It was expected that convergent validity would be demonstrated by significant within-category correlations. In addition, if within-category correlations were stronger than cross-category correlations, this would suggest that a particular informant or format demonstrated discriminant validity (Campbell & Fiske, 1959). Within-category correlations obtained from the same informant were calculated to evaluate the correspondence of adolescents and parents as informants of IA, HI, and CD. Within-category correlations for both adolescents and parents were highly significant. That is, adolescent rating data were significantly (ps $< .001$) correlated with adolescent interview data for IA ($r = .77$), HI ($r = .66$), and CD ($r = .76$). In addition parent rating data were significantly (ps $< .001$) correlated with parent interview data for IA ($r = .90$), HI ($r = .81$), and CD ($r = .54$). As recommended by Cohen and Cohen (1983), Fisher's z transformations were used to test the significance of the differences between independent correlations across informants. When informants were compared, adolescents demonstrated significantly higher correspondence for CD than did parents ($z = 3.92$, $p < .001$) and parents demonstrated significantly higher correspondence for IA ($z = 4.52$, $p < .001$) and HI ($z = 3.34$, $p < .001$) than did adolescents.

To examine discriminant validity of adolescent and parent reports for IA, HI, and CD, cross-category correlations (e.g., parent interview-IA with parent rating-CD) were compared to within-category correlations (e.g., parent interview-IA with parent rating-IA). It was expected that relevant cross-category correlations would be significantly weaker than within-category correlations, demonstrating adequate discrimination among traits (see Table V). T-tests were conducted to compare the within-category correlation to each cross-category correlation using the formula provided by Steiger (1980) for testing the significance of the difference between dependent correlations. Analyses revealed that the within-category correlation for adolescent reports of IA was significantly higher than all four of the relevant cross-category correlations (ps $< .001$). For adolescent reports of HI, the within-category correlation for adolescent HI was significantly higher than the cross-category correlations for three out of the four comparisons. For adolescent reports of CD, the within-category correlation for adolescent CD was significantly higher than all cross-category correlations. For parents, t-tests revealed that within-category

Table V. Multitrait-Multimethod Correlation Matrix

	Adolescent interviews			Parent interviews			Adolescent ratings			Parent ratings		
	IA	HI	CD	IA	HI	CD	IA	HI	CD	IA	HI	CD
Adolescent interviews												
IA	(0.82)											
HI	0.53*	(0.66)										
CD	0.30*	0.49*	(0.75)									
Parent interviews												
IA	$0.55^{a,*}$	$0.32^{b,*}$	0.13^{b}	(0.93)								
HI	$0.26^{b,*}$	0.19^{a}	0.13^{b}	0.62*	(0.84)							
CD	$0.27^{b,*}$	0.19^{b}	$0.35^{a,*}$	0.40*	0.39*	(0.55)						
Adolescent ratings												
IA	$0.77^{c,*}$	$0.51^{d,*}$	$0.29^{d,*}$	0.39*	0.08	0.17	(0.83)					
HI	$0.46^{d,*}$	$0.66^{c,*}$	$0.44^{d,*}$	0.20	0.18	0.17	0.55*	(0.67)				
CD	$0.27^{d,*}$	$0.43^{d,*}$	$0.76^{c,*}$	0.05	0.04	0.23*	0.27*	0.38*	(0.84)			
Parent ratings												
IA	0.48*	0.25*	0.13	$0.90^{c,*}$	$0.63^{d,*}$	$0.40^{d,*}$	$0.32^{a,*}$	0.16^{b}	0.03^{b}	(0.94)		
HI	0.16	0.19	0.14	$0.56^{d,*}$	$0.81^{c,*}$	$0.38^{d,*}$	0.02^{b}	0.14^{a}	0.07^{b}	0.62*	(0.86)	
CD	0.05	0.06	0.17	0.22^{d}	$0.39^{d,*}$	$0.54^{c,*}$	-0.04^{b}	0.02^{b}	0.12^{a}	0.25*	0.42*	(0.63)

Note. IA = inattention; HI = hyperactivity/impulsivity; CD = conduct disorder. Parentheses denote coefficient alpha values.
[a]Within-category, within-format correlation.
[b]Cross-category, within-format correlation.
[c]Within-category, within-informant correlation.
[d]Cross-category, within-informant correlation.
*$p < .001$, one-tailed.

correlations were significantly higher than all cross-category correlations (ps $< .001$) for IA and HI. For parent reports of CD, the within-category correlation for parent CD was significantly higher than only one of the four cross-category correlations.

Within-category correlations, obtained from different informants using the same format, were calculated to evaluate the correspondence of interviews and rating scales. Adolescent interview data were significantly (ps $< .001$) correlated with parent interview data for IA ($r = .55$) and CD ($r = .35$) but not for HI ($r = .19$). In addition, adolescent rating data were significantly ($p < .001$) correlated with parent rating data for IA ($r = .32$) but not for HI ($r = .14$) or CD ($r = .12$). Again, Fisher's z transformations were used to test the significance of the differences across formats. Interviews produced significantly higher (ps $< .001$) correspondence for IA ($z = 4.06$) and CD ($z = 3.46$) than did rating scales. There was no significant difference for HI. To examine discriminant validity of interview and rating scale reports of IA, HI, and CD, cross-category correlations (e.g., parent interview-IA with adolescent interview-CD) were compared to within-category correlations (e.g., parent interview-IA with adolescent interview-IA). For interviews, the within-category correlation for interviews of IA was significantly higher than all relevant cross-category correlations (ps $< .001$). However, for interviews of HI, the within-category correlation was not significantly higher than any of the cross-category correlations. For interviews of CD, the within-category correlation for interview reports of CD was significantly higher than two of the four cross-category correlations. For rating scales, the within-category correlations were significantly higher than three out of four cross-category correlations (ps $< .001$). For HI and CD, the within-category correlations were not significantly higher than any of the relevant cross-category correlations.

DISCUSSION

From this study conclusions can be drawn about the construct validity of three DBD traits and the utility of parent and adolescent reports, on the basis of interview and rating scale formats, for collecting information about DBD in adolescents. The CFA results provided evidence for construct validity of the traits of IA, HI, and CD. The model comparisons provided strong evidence for convergent and discriminant trait validities with little evidence for method effects. In addition, the examination of the factor correlations provided additional support for discriminant trait validity, as two of the three traits were not significantly correlated.

As predicted and consistent with previous studies, parent reports were more reliable than adolescent reports for IA and HI and adolescent reports were more reliable than parent reports for CD (Achenbach et al., 1987; Crowley et al., 2001; Loeber et al., 1989). Also consistent with predictions and previous studies, adolescent reports were more valid for CD than parent reports whereas parent reports were more valid for ADHD than adolescent reports (e.g., Andrews et al., 1993; Crowley et al., 2001). Adolescents may be better reporters of CD symptoms than their parents because these are discrete behaviors with low frequencies that may be more conspicuous during adolescence. Therefore, parents may not be privy to knowledge about these behaviors because they are not supervising their adolescents at all times. Parents may be better reporters of ADHD symptoms than adolescents since these behaviors are less discrete and may have been more conspicuous during childhood. Thus, they may be more evident to the observer than to the person who is displaying the behaviors. Therefore, adolescent self-reports should not be used exclusively for diagnosing ADHD nor should parent reports be used exclusively for diagnosing CD. This conclusion is consistent with previous recommendations to use multiple informants when assessing psychopathology in children and adolescents (e.g., Lahey et al., 1996).

Although it has been suggested that adult self-reports are reliable (Lahey et al., 1996), it may also improve diagnostic reliability to use multiple informants when diagnosing ADHD in adults. It is not clear from the current study whether the limited validity of self-reports of HI is due to age of client or the nature of the symptomatology. Stated differently, HI may be a trait that is more easily observed in others than in oneself. Future studies should examine whether adults are adequate self-reporters with regard to ADHD. The current findings suggest that adolescent self-reports of HI are less valid than reports from others (i.e., parents). Given that many college students are being evaluated for ADHD it is becoming increasingly important to determine whether reports from significant others (e.g., parent, roommate, or spouse) should routinely be collected.

With regard to formats, the findings were largely consistent with predictions. As expected, interviews were more reliable than ratings for IA and CD. However, the lack of difference in the reliability of interviews and ratings for HI was not predicted. Also consistent with predictions, interviews were more valid for IA and CD than were ratings. Again, the lack of difference in the validity of interviews and ratings for HI was not predicted. Thus, if time and resources allow, it is likely better to use semi-structured interviews than rating scales for assessing DBD. Semi-structured interviews may be better than ratings because expert clinical judgment is employed in making subjective decisions about the degree to which the behavior is maladaptive given the age and sex of adolescent and the degree to which the behavior causes impairment.

Although the current sample size was not large enough to analyze the data separately for boys and girls, sex has been implicated as a moderator variable in studies of interrater reliabilities (e.g., Epkins & Meyers, 1994; Seiffge-Krenke & Kollmar, 1998). In addition, a few significant sex by method interactions were found on the resulting behavioral symptoms counts. Therefore, future studies might examine differences in the reliability and validity of reports on the basis of sex of the children and/or adolescents. Future research might also explore the possibility of differential reliability and validity of parent reports on the basis of the parent sex. Research has suggested that mothers report more symptoms than fathers but these differences have not been shown to affect reliability (Christensen, Margolin, & Sullaway, 1992).

ACKNOWLEDGMENTS

This research was supported by grants from the University of Kentucky, Department of Psychiatry, to Drs. Martin and Milich; the Commonwealth Division of Mental Health Research Development Committee, the State Research Initiative, Kentucky Cabinet for Human Resources and a dissertation grant from NIMH to Dr. Hartung. We are grateful to Rick Hoyle, Don Lynam, Greg Smith, and James Grice for reviewing earlier versions of this manuscript and providing helpful comments. In addition, we thank the following individuals for their contributions to this project: Heather Cassidy, John Ireland, Justin Kyriannis, Caryn Letcher, Kelly McGraw, Rebecca Perkins, Shannon Sharp, Heather Turpin, and John Yozwiak.

REFERENCES

Achenbach, T. M., McConaughy, S. H., & Howell, C. T. (1987). Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychological Bulletin*, *101,* 213–232.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Andrews, V. C., Garrison, C. Z., Jackson, K. L., Addy, C. L., & McKeown, R. E. (1993). Mother–adolescent agreement on the symptoms and diagnoses of adolescent depression and conduct

disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, *32,* 731–738.

Bird, H. R., Gould, M. S., & Staghezza, B. (1992). Aggregating data from multiple informants in child psychiatry epidemiological research. *Journal of the American Academy of Child and Adolescent Psychiatry*, *31,* 78–85.

Byrne, B. M. (1994). Testing for construct validity: The multitrait-multimethod model. In *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56,* 81–105.

Cantwell, D. P., Lewinsohn, P. M., Rohde, P., & Seeley, J. R. (1997). Correspondence between adolescent report and parent report of psychiatric diagnostic data. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36,* 610–619.

Christensen, A., Margolin, G., & Sullaway, M. (1992). Interparental agreement on child behavior problems. *Psychological Assessment*, *4,* 419–425.

Cohen, J., & Cohen, P. (1983). Bivariate correlation and regression. In *Applied multiple regression/correlation analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crowley, T. J., Mikulich, S. K., Ehlers, K. M., Whitmore, E. A., & MacDonald, M. J. (2001). Validity of structured clinical evaluations in adolescents with conduct and substance problems. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40,* 265–273.

Edelbrock, C., Costello, A. J., Dulcan, M. K., Conover, N. C., & Kala, R. (1986). Parent–child agreement on child psychiatric symptoms assessed via structured interview. *Journal of Child Psychology and Psychiatry*, *27,* 181–190.

Epkins, C. C., & Meyers, A. W. (1994). Assessment of childhood depression, anxiety, and aggression: Convergent and discriminant validity of self-, parent-, teacher-, and peer-report measures. *Journal of Personality Assessment*, *62,* 364–381.

Frick, P. J., Silverthorn, P., & Evans, C. (1994). Assessment of childhood anxiety using structured interviews: Patterns of agreement among informants and association with maternal anxiety. *Psychological Assessment*, *6,* 372–379.

Greenbaum, P. E., Dedrick, R. F., Prange, M. E., & Friedman, R. M. (1994). Parent, teacher, and child ratings of problem behaviors of youngsters with serious emotional disturbance. *Psychological Assessment*, *6,* 141–148.

Hart, E. L., Lahey, B. B., Loeber, R., & Hanson, K. S. (1994). Criterion validity of informants in the diagnosis of disruptive behavior disorders in children: A preliminary study. *Journal of Consulting and Clinical Psychology*, *62,* 410–414.

Hartung, C. M., Milich, R., Lynam, D. R., & Martin, C. A. (2002). Understanding the relations among gender, disinhibition, and disruptive behavior in adolescents. *Journal of Abnormal Psychology*, *111,* 659–664.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Hu, L.-T., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112,* 351–362.

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, *112,* 165–172.

Lahey, B. B., Flagg, E. W., Bird, H. R., Schwab-Stone, M. E., Canino, G., Dulcan, M. K., et al. (1996). The NIMH methods for the epidemiology of child and adolescent mental disorders (MECA) study: Background and methodology. *Journal of the American Academy of Child and Adolescent Psychiatry*, *35,* 855–864.

Lahey, B. B., Schwab-Stone, M., Goodman, S. H., Waldman, I. D., Canino, G., Rathouz, P. J., et al. (2000). Age and gender differences in oppositional behavior and conduct problems: A cross-sectional household study of middle childhood and adolescence. *Journal of Abnormal Psychology*, *109,* 488–503.

Loeber, R., Green, S. M., Lahey, B. B., & Stouthamer-Loeber, M. (1989). Optimal informants of childhood disruptive behaviors. *Development and Psychopathology*, *1,* 317–337.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 177–198). Thousand Oaks, CA: Sage.

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.

Seiffge-Krenke, I., & Kollmar, F. (1998). Discrepancies between mothers' and fathers' perceptions of sons' and daughters' problem behaviour: A longitudinal analysis of parent–adolescent agreement on interalising and externalising problem behaviour. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *39,* 687–697.

Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, *87,* 245–251.

Thurber, S., & Snow, M. (1990). Assessment of adolescent psychopathology: Comparison of mother and daughter perspectives. *Journal of Clinical Child Psychology*, *19,* 249–253.

Tildesley, E. A., Hops, H., Ary, D., & Andrews, J. A. (1995). Multitrait-multimethod model of adolescent deviance, drug use, academic, and sexual behaviors. *Journal of Psychopathology and Behavioral Assessment*, *17,* 185–215.

Wechsler, D. (1981). *Manual for the Wechsler adult intelligence scale – Revised (WAIS-R)*. San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Manual for the Wechsler intelligence scale for children – Third edition (WISC-III)*. San Antonio, TX: The Psychological Corporation.

Widaman, K. F. (1985). Hierarchically tested covariance structure models for multitrait-multimethod data. *Applied Psychological Measurement*, *9,* 1–26.

Willcutt, E. G., Hartung, C. M., Lahey, B. B., Loney, J., & Pelham, W. E. (1999). Utility of behavior ratings by examiners during assessments of preschool children with Attention-Deficit/Hyperactivity Disorder. *Journal of Abnormal Child Psychology*, *27,* 463–472.