

RELIABILITY AND VALIDITY OF PARENT AND TEACHER RATINGS OF ATTENTION-DEFICIT/HYPERACTIVITY DISORDER SYMPTOMS

By: George J. DuPaul, Thomas J. Power, Kara E. McGoey, Martin J. Ikeda, Arthur D. Anastopoulos

DuPaul, G.J., Power, T.J., McGoey, K.E., Ikeda, M.J., & [Anastopoulos, A.D.](#) (1998). Reliability and validity of parent and teacher ratings of Attention-Deficit/Hyperactivity Disorder symptoms. *Journal of Psychoeducational Assessment*, 16, 55-68.

Made available courtesy of SAGE Publications (UK and US): <http://jpa.sagepub.com/>

*****Note: Figures may be missing from this format of the document**

Abstract:

The reliability and criterion-related validity of the Home and School versions of the AD/HD Rating Scale-IV were evaluated in a nonreferred sample of 71 students. Parent and teacher ratings were obtained 4 weeks apart at a time contemporaneous with observations of classroom behavior and academic productivity. Results indicated adequate levels of internal consistency, test-retest reliability, and cross-informant agreement for both parent and teacher ratings. Teacher ratings were significantly correlated with classroom observational data, and parent ratings were primarily related to behavior ratings. The discriminant validity of these scales also was examined in a sample of 92 clinic-referred children. Both the Home and School versions of the AD/HD Rating Scale-IV were found to discriminate significantly between children with and without AD/HD. The AD/HD Rating Scale-IV appears to have adequate psychometric properties for the screening and assessment of AD/HD.

Article:

Behavior rating scales completed by parents and teachers are among the most important measures used in assessing children and adolescents suspected of having Attention-Deficit/Hyperactivity Disorder (AD/HD) (Barkley, 1997; Hinshaw, 1994). Rating scales provide clinicians with valuable information regarding parent and teacher perceptions of the degree to which a child's AD/HD-related behaviors are deviant from peers of the same age and gender. Given the critical role that these questionnaires play in the assessment of AD/HD, it is imperative that such measures possess adequate levels of reliability and validity.

A plethora of behavior rating scales have been developed over the past two decades for use in assessing childhood behavior disorders. Broad-band measures, such as the Child Behavior Checklist (Achenbach, 1991a) and the Behavioral Assessment System for Children (Reynolds & Kamphaus, 1992), are used to obtain adult perceptions of a wide variety of possible behaviors (e.g., conduct problems, depression, anxiety, and social relationship difficulties). In addition, a number of narrow-band rating scales have been developed to assess behaviors specifically related to AD/HD. Some examples of these measures include the Attention Deficit Disorder Evaluation Scales (ADDES; McCarney, 1989), the ADD-H Comprehensive Teacher Rating Scale (ACTeRS; Ullmann, Sleator, & Sprague, 1985), the Disruptive Behavior Disorder Rating Scale (Pelham, Evans, Gnagy, & Greenslade, 1992), and the AD/HD Rating Scale (DuPaul,

1991). Although all of these narrow-band measures have been found to be reliable and valid, recent changes to the *Diagnostic and Statistical Manual of Mental Disorders (DSM-IV*, American Psychiatric Association, 1994) criteria for AD/HD have limited the utility of information derived from some questionnaires that were based on prior versions of the *DSM*. Further, other questionnaires that have been based on the DSM-IV criteria for AD/HD (e.g., AD/HD Symptom Checklist-4; Gadow & Sprafkin, 1997) lack extensive, representative normative data.

In response to changes in the *DSM* criteria for AD/HD, two versions of the AD/HD Rating Scale-IV (DuPaul, Anastopoulos, Power, Murphy, & Barkley, 1994) were constructed for use with parents and teachers. Both the Home and School versions of this rating scale require respondents to indicate the frequency of AD/HD symptoms on a 4-point Likert scale ranging from *never or rarely* (0) to *very often* (3). Normative data for a nationally representative sample of children and adolescents are available for both parent ratings (DuPaul, Anastopoulos, Power, Reid, Ikeda, & McGoey, in press) and teacher ratings (DuPaul, Power, Anastopoulos, Reid, McGoey, & Ikeda, 1997). These previous investigations have indicated that parent- and teacher-reported symptoms of AD/HD differ between student genders, as well as among age and ethnic groups. Further, ratings on the AD/HD Rating Scale-IV have been found to conform to a two-factor model (Inattention and Hyperactivity-Impulsivity) corresponding with the DSM-IV conceptualization of this disorder. Although these initial studies have provided support for the utility of this scale as a screening and diagnostic assessment instrument, psychometric properties of the AD/HD Rating Scale remain to be explicated.

The purpose of this investigation was to determine the internal consistency, test-retest reliability, interrater agreement, and criterion-related validity of both the Home and School versions of the AD/HD Rating Scale-IV. Parent and teacher ratings were obtained on two occasions (4 weeks apart) for a random sample of students from two suburban school districts. Criterion measures included parent and teacher ratings on the Conners Rating Scales (Conners, 1989), direct observations of classroom attention and activity level, as well as accuracy on classroom assignments. The discriminant validity of the AD/HD Rating Scale-IV was examined in a separate sample of children who had been referred to an outpatient clinic for an evaluation of AD/HD. Given the adequate psychometric properties associated with similar instruments, it was hypothesized that both versions of this questionnaire would provide scores that were internally consistent, stable, and significantly correlated with other measures of AD/HD-related behaviors. Further, it was predicted that scores on the AD/HD Rating Scale-IV would be significantly higher (indicating more frequent exhibition of AD/HD symptoms) for children diagnosed with AD/HD than for clinically referred children who did not receive a diagnosis of AD/HD.

METHOD

Participants

Two separate samples of children participated in this study. Sample 1 consisted of 71 students (35 boys, 36 girls) ranging in age from 5 to 17 years ($M = 11.0$; $SD = 3.4$) who were randomly selected from two suburban school districts located in eastern Pennsylvania and western New Jersey. Participants were predominantly Caucasian ($n = 60$) but also included children of African American ($n = 5$), Latino ($n = 4$), and Asian American ($n = 2$) backgrounds. All participants were placed in general education classrooms. This sample was used to examine the test-retest

reliability, internal consistency, and criterion-related validity of AD/HD ratings. Data regarding the ethnicity, sex, and years of teaching experience for teacher respondents were not available.

Teacher ratings. Complete data for analyzing the test-retest reliability of teacher ratings were available for 52 children (24 boys, 28 girls) from Sample 1 who ranged in age from 5 to 17 years ($M = 11.3$; $SD = 3.6$) and attended kindergarten through 12th grade ($M = 5.7$; $SD = 3.7$). For the validity analyses of teacher ratings, three of the criterion measures involved direct observation of classroom behavior for a subsample of 53 students (25 boys, 28 girls) who ranged in age from 5 to 14 years ($M = 9.8$; $SD = 2.6$) and attended kindergarten through 8th grade ($M = 4.0$; $SD = 2.6$).

Parent ratings. Complete test-retest reliability data for parent ratings were available for 43 children (17 boys, 26 girls) from Sample 1 who ranged in age from 5 to 17 years ($M = 11.07$; $SD = 3.50$) and attended kindergarten through 12th grade ($M = 5.35$; $SD = 3.59$). For the validity analyses, three of the criterion measures involved direct observation of classroom behavior for a subsample of 46 students (22 boys, 24 girls) who ranged in age from 5 to 14 years ($M = 10.09$; $SD = 2.58$) and attended kindergarten through 8th grade ($M = 4.30$; $SD = 2.62$).

Interobserver agreement. For analysis of interobserver agreement, parent and teacher ratings were available for a sample of 62 students (28 boys, 34 girls) from Sample 1. These students ranged in age from 5 to 17 years ($M = 11.1$; $SD = 3.4$) and attended kindergarten through 12th grade ($M = 5.4$; $SD = 3.5$).

Discriminant validity sample. Sample 2 consisted of consecutive referrals to the AD/HD Evaluation and Treatment Program of a regional pediatric hospital located in a large metropolitan area in the Northeastern region of the United States. This sample was included to examine the discriminant validity of parent and teacher AD/HD ratings. Children in this sample were referred for an initial evaluation or a reevaluation of AD/HD. Participants met the following inclusion criteria: (a) completion by parents and teachers of the AD/HD Rating Scale-IV and a diagnostic interview with parents using the Diagnostic Interview for Children and Adolescents-Revised (DICA-R; Reich, Shayka, & Taibleson, 1991); and (b) estimated IQ of 80 or above on the Kaufman Brief Intelligence Test (KBIT; Kaufman & Kaufman, 1990). Children were excluded if they presented with evidence of pervasive developmental disorder, a psychotic disorder, or a progressive neurological disorder. Also, children were excluded if they had been prescribed a psychotropic medication for AD/HD or related disorders within 6 months of the time of evaluation.

Sample 2 consisted of 92 children (24 girls, 68 boys) between the ages of 6-0 and 149 years ($M = 9.0$, $SD = 2.2$). Grade levels ranged from kindergarten through eighth grade, with 73% of the sample attending grades 1 through 4. The distribution of ethnic groups represented was 21.7% African American, 3.3% Latino/Hispanic, and 75% Caucasian. The range of socioeconomic levels as assessed by the Four Factor Index of Social Status (Hollingshead, 1975) was as follows: 3.2% in Category I (unskilled laborers), 14.2% in Category II (machine operators, semiskilled workers), 25% in Category III (skilled craftsman, clerical, sales workers), 40.2% in Category IV (small business owners, technicians), and 17.4% in Category V (major business owners, professionals). On the KBIT, the sample achieved mean scores of 103.1 ($SD = 11.9$) on the Vocabulary scale, 100.5 ($SD = 11.7$) on Matrices, and 101.9 ($SD = 11.1$) on the Composite.

Measures

Parents and teachers of children in both samples were asked to provide information about the child being rated, such as gender, grade, and age. For children in Sample 1, two behavior rating scales were completed by parents and teachers, and behavior observations were conducted in the classrooms of participating students.

AD/HD Rating Scale-N (School Version). Teachers of children in both samples completed the AD/HD Rating Scale-IV-School Version (DuPaul et al., 1994), which consists of 18 items directly adapted from the AD/HD symptom list as specified in the *DSM-IV*. In order to address possible response bias, Inattention symptoms were designated as odd-numbered items, while Hyperactivity-Impulsivity symptoms were displayed as even-numbered items. Teachers selected the single response for each item that best described the frequency of the specific behavior displayed by the target child over the past 6 months (or since the beginning of the school year). The frequency of each item or symptom was delineated on a 4-point Likert scale ranging from *never or rarely* (0) to *very often* (3), with higher scores indicative of greater AD/HD-related behavior. Factor analyses of the AD/HD Rating Scale-IV-School Version indicate that it is comprised of a 9-item Inattention factor and a 9-item Hyperactivity-Impulsivity factor (DuPaul, Power, Anastopoulos, Reid, McGoey, & Ikeda, 1997). Thus, these two factor scores as well as a total scale score were used as measures in the present study.

Conners Teacher Rating Scale-39. The Conners Teacher Rating Scale-39 (CTRS39; Conners, 1989) was completed by each teacher of children in Sample 1 for use as a criterion measure. The CTRS-39 contains 39 items wherein frequency of behavior is rated on a 4-point Likert scale ranging from *not at all* (0) to *very much* (3). This rating scale has adequate psychometric properties and has been widely used for clinical and research purposes with the AD/HD population (Barkley, 1990).

AD/HD Rating Scale-IV (Home Version). Parents of children in both samples completed the AD/HD Rating Scale-IV-Home Version (DuPaul et al., 1994), which consists of 18 items directly adapted from the AD/HD symptom list as specified in the *DSM-IV*. Parents selected the single response for each item that best described the frequency of the specific behavior displayed by the target child over the past 6 months. In order to address possible response bias, Inattention symptoms were designated as odd-numbered items and Hyperactivity-Impulsivity symptoms were displayed as even-numbered items. The frequency of each item or symptom was delineated on a 4-point Likert scale ranging from *never or rarely* (0) to *very often* (3), with higher scores indicative of greater AD/HD-related behavior. As was the case for the school version of this scale, the AD/HD Rating Scale-IV-Home Version has been found to be comprised of a 9-item Inattention factor and a 9-item Hyperactivity-Impulsivity factor (DuPaul et al., in press). These two scale scores as well as a total scale score were used in the present study.

Conners Parent Rating Scale-48. Parents of children in Sample 1 also completed the Conners Parent Rating Scale-48 (CPRS-48; Conners, 1989) as a criterion measure. The CPRS-48 contains 48 items wherein the frequency of each item is rated on a 4-point Likert scale ranging from *not at all* (0) to *very much* (3). The CPRS-48 has adequate psychometric properties and is widely used for clinical and research purposes with the AD/HD population (Barkley, 1990).

Behavioral observations. The classroom behavior of children in Sample 1 who were placed in kindergarten through eighth grade was observed by one or more of the investigators using an adaptation of the AD/HD Behavior Code originally designed by Barkley (1990). The occurrence of two behaviors (i.e., Off-task and Fidgets) was recorded on a partial interval basis using observation intervals of 10 sec, with 5 sec between intervals used for recording observed behaviors. Off-task behavior was defined as the student breaking eye contact with task materials or classroom instruction for at least 3 consecutive seconds. Fidgets was defined as purposeless motion of the legs, arms, hands, buttocks, or trunk that occurred at least four times in succession. Each observation session was approximately 15 min in length. For each behavior, the percentage of intervals where the behavior occurred was calculated by dividing the number of intervals in which the behavior occurred by the total number of observation intervals and multiplying the dividend by 100%.

An academic efficiency score (AES; Rapport, DuPaul, & Kelly, 1989) was calculated for each participant who was observed in the classroom. For each student, teachers submitted three samples of independent seatwork completed in the classroom. The specific content of independent seatwork varied across participants and typically included written math problems and language workbook assignments. Although the nature of this seatwork was not controlled for in this experiment, the use of teacher-derived assignments was believed to enhance the ecological validity of this measure. The AES was calculated by dividing the number of work items completed correctly by the number of items assigned and multiplying the dividend by 100%. This score represented the quality of each child's academic performance in relation to his or her classmates, who were presumably asked to complete the same amount of work at similar levels of difficulty.

Procedures

Parent and teacher ratings for Sample 1 were obtained over a 1-month period in May and June of 1995 (to ensure teacher familiarity with student behavior). Within each of the two districts, two to four children (with an attempt to include equivalent numbers of boys and girls) at each grade level were randomly selected to participate. Written, informed consent was then requested from each student's parent(s). If parent permission was not granted, then another child (of the same gender) was randomly selected and permission was requested from that child's parent(s). No teachers provided ratings for more than two children in their classrooms. Only children in general education classrooms participated.

Once written parent permission was obtained, parents and teachers were asked to complete the appropriate version of the AD/HD Rating Scale-IV on two occasions 4 weeks apart to assess test-retest reliability. On each occasion, ratings were completed on Fridays and were to reflect observations of the child's behavior over the previous week. The CTRS-39 was also completed by the teacher on one of the two occasions, and the CPRS-48 was completed by the parent on one of the two occasions. For all children in grades K-8, one of the investigators conducted behavioral observations on three separate days (i.e., a total of 45 min of observation) during one of the two weeks when parents and teachers were due to complete the AD/HD Rating Scale-IV. The observer was situated in a part of the classroom that avoided direct eye contact with the target child, but at a distance that allowed easy determination of on-task and fidgety behavior.

Following each observation, the teacher provided information necessary to calculate an AES score (i.e., how much work was completed correctly relative to the amount assigned). A second observer was present for 30% of the observation sessions so that interobserver agreement could be determined. Interobserver agreement was calculated by dividing the number of agreements by the total number of agreements plus disagreements and multiplying the dividend by 100%. Agreement was consistently above 80% and averaged 88% across the two behavioral categories. A mean kappa coefficient of .56 was obtained to demonstrate agreement beyond chance levels.

Children in Sample 2 were assigned to a diagnostic group or clinical control group based on their scores on a multimethod assessment battery including the parent version of the DICA-R, the parent-rated Child Behavior Checklist (CBCL; Achenbach, 1991b), and the teacher-rated Child Attention Problems scale (CAP; Barkley, 1990). Children were categorized as having AD/HD Predominantly Inattentive Type (AD/HD-I) if they demonstrated the following: (a) DICA-R diagnosis of AD/HD-I; (b) a *T* score of 60 or above on the Attention Problems factor of the CBCL; and (c) a score on the Inattention subscale of the CAP of greater than or equal to the 93rd percentile. Children were diagnosed with AD/HD Combined Type (AD/HD-COM) if they demonstrated: (a) DICA-R diagnosis of AD/HD-COM; (b) a *T* score of 60 or above on the Attention Problems factor of the CBCL; and (c) scores on the Inattention and Overactivity subscales of the CAP of greater than or equal to the 93rd percentile. Two children met criteria for AD/HD Predominantly Hyperactive Impulsive type and were not included in further analyses. Remaining children who did not meet criteria for any of the AD/HD subtypes were assigned to a clinical control group.

Based on these criteria, 30 children were classified as having AD/HD-I, 25 participants had AD/HD-COM, and 35 children were assigned to the clinical control group. Although the AD/HD-COM group had a higher proportion of participants with a comorbid conduct disorder, there were no further differences between groups with respect to psychiatric comorbidity, gender, age, or special education placement.

Parents and teachers of participants in Sample 2 completed the CBCL, CAP, and Home and School versions of the AD/HD Rating Scale-IV prior to their initial clinic visit. The DICA-R was conducted by a doctoral-level psychology clinician or an advanced doctoral candidate in psychology at the initial clinic visit. Administrations of the DICA-R were audiotaped; 23% of the audiotapes were selected at random and reviewed by another clinician to establish inter-scoring reliability. Kappa coefficients were .87 for a diagnosis of AD/HD-COM, .81 for AD/HD-I, .91 for oppositional defiant disorder, .78 for conduct disorder, .84 for anxiety disorders, and .65 for mood disorders; all kappa coefficients are above the established limits of acceptability (Hartman, 1982).

RESULTS

Internal Consistency, Reliability, and Interobserver Agreement for AD/HD Ratings

Coefficient alphas were calculated to determine the internal consistency of the AD/HD Rating Scale-IV-School Version and its two subscales. The following alpha coefficients were obtained: Total Score = .94, Inattention (IA) = .96, and Hyperactivity-Impulsivity (HI) = .88. Test-retest reliability data were obtained for teacher ratings occurring 4 weeks apart. Pearson product-

moment correlation coefficients were as follows: Total Score = .90, Inattention = .89, and Hyperactivity-Impulsivity = .88.

In similar fashion, coefficient alphas were calculated to determine the internal consistency of the AD/HD Rating Scale-IV-Home Version and its two subscales. The following alpha coefficients were obtained: Total Score = .92, Inattention = .86, and Hyperactivity-Impulsivity = .88. Test-retest reliability data were obtained for parent ratings occurring 4 weeks apart. Pearson product-moment correlation coefficients were as follows: Total Score = .85, Inattention = .78, and Hyperactivity-Impulsivity = .86. Interrater agreement coefficients between parents and teachers were in the moderate range, as follows: Total Score = .41, Inattention = .45, and Hyperactivity-Impulsivity = .40.

Relationships among Teacher AD/HD Ratings and Criterion Measures

Pearson product-moment correlations between AD/HD Rating Scale-IV-School Version scores and criterion measures (i.e., CTRS-39 scores, direct observations of off-task and fidgety behavior, and mean AES) are presented in Table 1. Overall, the absolute values of obtained Pearson correlation coefficients ranged from .22 to .88, with 28 out of 30 achieving statistical significance. If a Bonferroni correction is applied to control Type 1 error rate for multiple correlations (i.e., $\alpha = .002$), 20 of these correlations can be considered statistically significant. As expected, the strongest correlations were found between AD/HD Rating Scale-IV-School Version factor scores and CTRS-39 Hyperactivity and Hyperactivity Index scores. In fact, ratings of AD/HD symptoms on these two measures shared between 53% and 77% of the variance. The correlation between the HI subscale and CTRS-39 Conduct Problems was significantly greater than the correlation between the IA subscale and Conduct Problems, $t(70) = 2.54, p < .01$. Conversely, the correlations between the IA subscale and CTRS-39 Anxious-Passive and Daydreams-Attention scales were significantly greater than the correlations between these two CTRS-39 scales and the HI subscale, $t(70) = 1.99, p < .05$, and $t(70) = 6.13, p < .001$, respectively. No other significant differences between the IA and HI subscales were obtained.

Table 1
Validity Coefficients for AD/HD Rating Scale-IV (School Version)

Measure	Inattention	Hyperactivity-Impulsivity	Total
CTRS Hyperactivity ^a	.73***	.79***	.86***
CTRS Conduct Problems ^a	.29*	.55***	.44***
CTRS Emot-Indulgence ^a	.54***	.41**	.56***
CTRS Anxious-Passive ^a	.47***	.25*	.45***
CTRS Asocial ^a	.43***	.36***	.46***
CTRS Daydream-Att ^a	.85***	.44***	.80***
CTRS Hyper Index ^a	.76***	.76***	.88***
Mean Off-Task ^b	.35**	.22	.34**
Mean Fidgets ^b	.28*	.23	.29*
Mean Accuracy ^b	-.46***	-.34**	-.47***

Note.—CTRS = Conners Teacher Rating Scale-39; Emot-Indulgence = Emotional Indulgence; Daydream-Att = Daydream-Attention Problems; Hyper Index = Hyperactivity Index.

^a $n = 71$. ^b $n = 53$.

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

The AD/HD Rating Scale-P/-School Version Total Score and IA subscale score were significantly correlated with direct observations of classroom off- task and fidgety behavior (see Table 1). The correlations between the HI subscale and classroom behavior were nonsignificant. All three AD/HD Rating Scale-IV-School Version scores were significantly associated, in a negative fashion, with accuracy on academic tasks. It should be noted that correlations with classroom observation measures were lower than correlations with CTRS-39 ratings. In fact, only two of the nine correlations with observational measures were significant at the .002 alpha level. Higher teacher ratings of AD/HD symptoms were associated with lower levels of task accuracy and, in the case of the IA subscale, higher frequencies of off-task and fidgety behavior.

Relationships among Parent AD/HD Ratings and Criterion Measures

Pearson product-moment correlations between AD/HD Rating Scale-IV-Home Version scores and CPRS-R scores are presented in Table 2. Overall, the absolute values of obtained validity coefficients ranged from .10 to .81, with 15 out of 18 achieving statistical significance. If a Bonferroni correction is applied to control Type 1 error rate for multiple correlations (i.e., $\alpha = .003$), 12 of these correlations can be considered statistically significant. As expected, the strongest correlations were found between AD/HD Rating Scale-IV-Home Version and CPRS-48 Hyperactivity Index scores. In fact, ratings of AD/HD symptoms on these two measures shared between 37% and 66% of the variance. The pattern of correlations provided initial evidence for the discriminant validity of the IA and HI subscales. Significantly stronger correlations were obtained between the HI subscale and the CPRS-48 Conduct Problems, $t(56) = 2.19, p < .05$, Impulsive-Hyperactive, $t(56) = 4.65, p < .001$, and Hyperactivity Index, $t(56) = 2.99, p < .01$, scores than were found for the IA subscale and these three indices. Conversely, the IA subscale was more strongly correlated with the CPRS-48 Learning Problems scale than was the HI subscale, $t(56) = 2.44, p < .01$. Neither AD/HD Rating Scale-IV subscale correlated significantly with CPRS-48 Anxious ratings.

Parent ratings of AD/HD symptoms were significantly correlated with teacher ratings on the CTRS-39 Hyperactivity and Conduct Problems factors as

Table 2
Correlation Coefficients between Parent Ratings on the AD/HD Rating Scale-IV-Home Version and Conners Parent Rating Scale-Revised

Measure	Inattention	Hyperactivity-Impulsivity	Total
CPRS Conduct Problems	.45***	.65***	.61***
CPRS Learning Problems	.66***	.45***	.60***
CPRS Psycho	.28*	.32**	.36**
CPRS Imp-hyp	.45***	.78**	.68***
CPRS Anxious	.05	.18	.10
CPRS Hyper Index	.61***	.81***	.80***

Note.— $n = 43$. CPRS = Conners Parent Rating Scale-48.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 3
Correlations between Parent Ratings on the AD/HD Rating Scale-IV-Home Version and School Behavior

Measure	Inattention	Hyperactivity-Impulsivity	Total
CTRS Anxious-Passive ^a	.18	.14	.20
CTRS Asocial ^a	.13	.14	.20
CTRS Conduct Prob ^a	.39**	.26*	.38**
CTRS Daydreams-Att ^a	.32*	.18	.28*
CTRS Emotional Ind ^a	.15	-.02	.12
CTRS Hyperactivity ^a	.38**	.35**	.41**
CTRS Hyper Index ^a	.30*	.31*	.34*
Mean Off-Task ^b	.28	.14	.26
Mean Fidgets ^b	.15	.17	.25
Mean Accuracy ^b	-.43**	-.18	-.36**

Note.—CTRS = Conners Teacher Rating Scale-39.

^a*n* = 62. ^b*n* = 46.

p* < .05. *p* < .01.

well as the Hyperactivity Index (see Table 3). In addition, parent IA and Total scores were correlated with the CTRS-R Daydreams-Attention Problem factor. AD/HD Rating Scale-IV-Home Version scores were not correlated with teacher ratings of anxious-passive, asocial, or emotionally indulgent behavior. Contrary to expectations, parent ratings were not significantly correlated with classroom observations of either off-task or fidgety behavior (see Table 3). Alternatively, both the IA subscale and Total Score were significantly associated, in a negative fashion, with accuracy on academic tasks. Thus, higher parent ratings of Inattentive symptoms and Total Score were associated with lower levels of task accuracy. It should be noted that the correlations between parent ratings and school validity data were relatively low and, with a Bonferroni correction ($\alpha = .002$), none of the obtained coefficients reached statistical significance.

Discriminant Validity of Parent and Teacher Ratings

Means and standard deviations for parent and teacher IA and HI scores across three groups (AD/HD-COM, AD/HD-I, and Clinical Control) are presented in Table 4. Statistically significant differences in mean ratings among these three groups were obtained for parent IA ratings, $F(2, 87) = 7.56, p < .001$, parent HI ratings, $F(2, 87) = 5.60, p < .01$, teacher IA ratings, $F(2, 87) = 22.34, p < .0001$, and teacher HI ratings, $F(2, 87) = 23.57, p < .0001$. Tukey HSD post hoc comparisons (conducted using an alpha level of .05) indicated that parent and teacher IA ratings were significantly higher (indicating greater levels of inattention) for participants in the predominantly inattentive and combined- type groups relative to clinical controls. Alternatively, parent and teacher HI ratings were significantly higher for participants in the combined-type group relative to their counterparts in the other two groups. There were no significant differences in parent and teacher HI scores between the predominantly inattentive type and clinical control participants.

Table 4
Means and Standard Deviations of AD/HD Rating Scale-IV Scores across Clinical Groups

Measure	Group		
	Control	AD/HD-I	AD/HD-COM
Parent Inattention	14.2 (7.9) ^a	19.3 (4.4) ^b	19.3 (4.3) ^b
Parent Hyperactivity-Impulsivity	11.6 (8.0) ^a	10.7 (5.7) ^a	16.4 (5.9) ^b
Teacher Inattention	13.3 (5.9) ^a	19.3 (4.7) ^b	21.6 (4.3) ^b
Teacher Hyperactivity-Impulsivity	10.5 (8.0) ^a	6.9 (4.5) ^a	18.6 (5.7) ^b

Note.—AD/HD-I = AD/HD Predominantly Inattentive Type; AD/HD-COM = AD/HD Combined Type. Standard deviations are in parentheses.

DISCUSSION

Teacher and parent ratings of AD/HD symptoms, as defined in the *DSM-IV*, were found to be internally consistent and stable over a 4-week period. The reliability of the Hyperactivity-Impulsivity scale was similar for the School and Home versions, but the internal consistency and stability of the Inattention scale were stronger for the School Version. This latter finding suggests that teachers may be somewhat more reliable than parents in their reporting of information about a child's inattentive behavior using items from the *DSM-IV*.

Interrater agreement between parents and teachers was in the moderate range, which is consistent with prior research using the Child Behavior Checklist and Teacher Report Form (Achenbach, 1991a), the Conners Parent and Teacher Rating Scales (Conners, 1989), and the previous version of the AD/HD Rating Scale (DuPaul, 1991). The relatively low amount of shared variance between the Home and School versions suggests that parents and teachers contribute unique information to the understanding of children's inattentive and hyperactive-impulsive behavior.

The pattern of correlations between factors of the AD/HD Rating Scale-IV and criterion measures provided some support for the discriminant validity of the Inattention and Hyperactivity-Impulsivity subscales. As expected, the Hyperactivity-Impulsivity factor demonstrated a much stronger correlation with the Conners Conduct Problems subscale than did the Inattention factor (Lahey, Schaughency, Hynd, Carlson, & Neives, 1987); this finding emerged for both parent and teacher ratings. The finding that parent ratings on the Conners Impulsive-Hyperactive factor were more strongly correlated with the Hyperactivity-Impulsivity factor than the Inattention factor was also anticipated. The similar correlations between the Conners Hyperactivity scale and the two factors of the AD/HD Rating Scale-IV (i.e., .73 for Inattention, .79 for Hyperactivity-Impulsivity) on the School Version most likely was due to the heterogeneity of items on the Conners subscale, including items referring to hyperactivity (restless or overactive) and inattention (inattentive, easily distracted).

Consistent with a two-factor model for AD/HD, teacher ratings on the Inattention factor demonstrated a much stronger relationship with the Conners Daydreams-Attention Problems scale than did teacher ratings on the Hyperactivity-Impulsivity factor. Similarly, Inattention scores derived from the Home Version were more strongly correlated with the Conners Learning Problems scale, which is comprised mainly of inattention items, than were Hyperactivity-Impulsivity scores. The relationship between the Inattention factor and ratings on Conners factors pertaining to anxiety was significant for teachers but not for parents. Given prior research

that has failed to demonstrate a consistent relationship between attention problems and anxiety (Achenbach, 1991b, 1991c; Conners, 1989), additional research is needed to explicate the meaning of this variation.

Behavioral observations of off-task behavior and academic accuracy during seatwork were related to the Inattention factor, but correlations between the Hyperactivity-Impulsivity factor and observational measures generally were not significant. In particular, the correlation between the Hyperactivity-Impulsivity factor and observations of fidgeting behavior was lower than expected and suggests a lack of correspondence between repetitive fidgets, as defined in Barkley (1991), and the symptoms pertaining to Hyperactivity-Impulsivity as delineated in *DSM-IV*. Also, the Hyperactivity-Impulsivity factor might have been more highly correlated with classroom observations if the observational code had been broader in scope (i.e., included categories specific to impulsive behavior).

As expected, teacher ratings of AD/HD symptoms appeared to have a stronger relationship with direct observations of off-task behavior, fidgeting, and work accuracy than did parent ratings. This finding suggested that teacher ratings of AD/HD symptoms are more indicative of students' behavior in school than are parent ratings. Thus, when clinicians are unable to conduct a direct observation of behavior, teacher ratings are preferable to parent ratings in determining the nature and extent of school problems related to AD/HD.

An initial examination of the discriminant validity of the AD/HD Rating Scale-IV indicated that both parent and teacher ratings on this measure reliably discriminated between children diagnosed with AD/HD and clinical controls. Consistent with the factor structure of this scale, parent and teacher ratings of hyperactivity-impulsivity symptoms were significantly greater for children diagnosed with the combined type of AD/HD relative to the remaining two groups. Stated differently, the HI factor was found to have discriminant validity with respect to differentiating between two subtypes of AD/HD as well as between clinical controls and children with the combined type of AD/HD.

A limitation of this research is that it was conducted in school districts and a clinic setting in the northeastern part of the United States. Although there is no reason to think that findings pertaining to the reliability and validity of the AD/HD Rating Scale-IV would be appreciably different in other regions of the country, this possibility needs to be examined. Also, certain ethnic minority groups, in particular African Americans, were somewhat underrepresented in Sample 1, resulting in questions about the applicability of the findings to these groups. Further, the low number of participants in each minority group did not permit an analysis of the data by subgroup, which is necessary to demonstrate the applicability of the findings for that group. As a result, the psychometric properties of the AD/HD Rating Scale-IV for each minority group need to be demonstrated so that this measure can be used with confidence with each population.

Another limitation of the study is that the range of criterion variables, particularly those measured by direct observation, was somewhat narrow. For instance, useful criterion measures for Hyperactivity-Impulsivity may have included direct observations of out-of-seat and calling-out behaviors. Also, a desk check for organization (Atkins, Pelham, & Licht, 1985) may have provided a useful criterion measure for the Inattention factor.

In conclusion, this study supports the internal consistency and stability of parent and teacher ratings of DSM-IV-symptoms pertaining to Inattention and Hyperactivity-Impulsivity. In addition, the findings provided some confirmation to the two-dimensional structure of AD/HD outlined in the *DSM-IV*. The Hyperactivity-Impulsivity factor was more related to conduct problems than was the Inattention factor. The relationships between teacher ratings of Inattention and observations of off-task behavior and work accuracy were significant, but not stronger than the correlations between teacher ratings of Hyperactivity-Impulsivity and direct observation measures. Further, the two factors of the AD/HD Rating Scale-IV can be used to differentiate children with AD/HD from other clinic-referred children as well as differentiate among the combined and inattentive subtypes of AD/HD. The AD/HD Rating Scale-IV appears to have acceptable reliability and validity for clinical use as a screening and diagnostic instrument particularly when used in the context of a multi-method assessment battery that includes diagnostic interviews, multiple behavior rating scales, and behavioral observations. Additional research is needed to verify the generalizability of the properties and utility of this scale across geographic regions, ethnic groups, and socioeconomic groups comprising the U.S. population.

REFERENCES

- Achenbach, T. M. (1991a). *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF Profiles*. University of Vermont: Department of Psychiatry.
- Achenbach, T. M. (1991b). *Manual for the Child Behavior Checklist/4-18 and 1991 Profile*. University of Vermont: Department of Psychiatry.
- Achenbach, T. M. (1991c). *Manual for the Teacher's Report Form and 1991 Profile*. University of Vermont: Department of Psychiatry.
- American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Atkins, M. S., Pelham, W. E., & Licht, M. H. (1985). A comparison of objective classroom measures and teacher ratings of attention deficit disorder. *Journal of Abnormal Child Psychology*, *13*, 155-167.
- Barkley, R. A. (1990). *Attention Deficit Hyperactivity Disorder: A handbook for diagnosis and treatment*. New York: Guilford Press.
- Barkley, R. A. (1991). *Attention Deficit Hyperactivity Disorder: A clinical workbook*. New York: Guilford Press.
- Barkley, R. A. (1997). Behavioral inhibition, sustained attention, and executive functions: Constructing a unifying theory of AD/HD. *Psychological Bulletin*, *121*, 65-94.
- Conners, C. K. (1989). *Conners Rating Scales Manual*. North Tonawanda, NY: Multi-Health Systems.
- DuPaul, G. J. (1991). Parent and teacher ratings of AD/HD symptoms: Psychometric properties in a community-based sample. *Journal of Clinical Child Psychology*, *20*, 245-253.
- DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Murphy, K., & Barkley, R. A. (1994). *AD/HD Rating Scale-IV*. Unpublished rating scale.
- DuPaul, G. J., Anastopoulos, A. D., Power, T. J., Reid, R., Ikeda, M., & McGoey, K. (in press). Parent ratings of attention deficit/hyperactivity disorder symptoms: Factor structure and normative data. *Journal of Psychopathology and Behavioral Assessment*.

- DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Reid, R., McGoey, K., & Ikeda, M. (1997). Teacher ratings of attention-deficit/hyperactivity disorder: Factor structure and normative data. *Psychological Assessment, 9*, 436-444.
- Gadow, K., & Sprafkin, J. (1997). *AD/HD Symptom Checklist-4*. Stony Brook, NY: Checkmate Plus.
- Hartman, D. P. (1982). *Using observers to study behavior: New directions for methodology of social and behavioral science*. San Francisco: Jossey-Bass.
- Hinshaw, S. P. (1994). *Attention deficits and hyperactivity in children*. Thousand Oaks, CA: Sage.
- Hollingshead, A. B. (1975). *Four Factor Index of Social Status*. New Haven, CT: Yale University.
- Kaufman, A. S., & Kaufman, N. L. (1990). *Kaufman Brief Intelligence Test*. Circle Pines, MN: American Guidance Service.
- Lahey, B. B., Schaughency, E. A., Hynd, G. W., Carlson, C. L., & Neives, N. (1987). Attention deficit disorder with and without hyperactivity: Comparison of behavioral characteristics of clinic-referred children. *Journal of the American Academy of Child and Adolescent Psychiatry, 26*, 718-723.
- McCarney, S. B. (1989). *The Attention Deficit Disorders Evaluation Scale: School version technical manual*. Columbia MO: Hawthorne Educational Services.
- Pelham, W. E., Jr., Evans, S. W., Gnagy, E. N., & Greenslade, K. E. (1992). Teacher ratings of DSM-III-R symptoms for the disruptive behavior disorders: Prevalence, factor analyses, and conditional probabilities in a special education sample. *School Psychology Review, 21*, 285-299.
- Rappport, M. D., DuPaul, G. J., & Kelly, K. L. (1989). Attention-Deficit Hyperactivity Disorder and methylphenidate: The relationship between gross body weight and drug response in children. *Psychopharmacology Bulletin, 25*, 285-290.
- Reich, W., Shayka, M. A., & Taibleson, C. (1991). *Diagnostic Interview for Children and Adolescents—DSM III-R Version (Parent Form)*. St. Louis, MO: Washington University Division of Child Psychiatry.
- Reynolds, C. R., & Kamphaus, R. (1992). *Behavior assessment system for children (BASC)*. Circle Pines, MN: American Guidance Service.
- Ullmann, R. K., Sleator, E. K., & Sprague, R. L. (1985). Introduction to the use of the ACTeRS. *Psychopharmacology Bulletin, 21*, 915-920.